The Handbook of Linguistics

Blackwell Handbooks in Linguistics

This outstanding multi-volume series covers all the major subdisciplines within linguistics today and, when complete, will offer a comprehensive survey of linguistics as a whole.

The Handbook of Child Language
Edited by Paul Fletcher and Brian MacWhinney

The Handbook of Phonological Theory, Second Edition
Edited by John A. Goldsmith, Jason Riggle, and Alan C. L. Yu

The Handbook of Contemporary Semantic Theory
Edited by Shalom Lappin

The Handbook of Sociolinguistics
Edited by Florian Coulmas

The Handbook of Phonetic Sciences, Second Edition
Edited by William J. Hardcastle and John Laver

The Handbook of Morphology
Edited by Andrew Spencer and Arnold Zwicky

The Handbook of Japanese Linguistics
Edited by Natsuko Tsujimura

The Handbook of Linguistics, First Edition
Edited by Mark Aronoff and Janie Rees-Miller

The Handbook of Contemporary Syntactic Theory
Edited by Mark Baltin and Chris Collins

The Handbook of Discourse Analysis
Edited by Deborah Schiffrin, Deborah Tannen, and Heidi E. Hamilton

The Handbook of Language Variation and Change, Second Edition
Edited by J. K. Chambers and Natalie Schilling

The Handbook of Historical Linguistics
Edited by Brian D. Joseph and Richard D. Janda

The Handbook of Language, Gender, and Sexuality, Second Edition
Edited by Susan Ehrlich, Miriam Meyerhoff, and Janet Holmes

The Handbook of Second Language Acquisition
Edited by Catherine J. Doughty and Michael H. Long

The Handbook of Bilingualism and Multilingualism, Second Edition
Edited by Tej K. Bhatia and William C. Ritchie

The Handbook of Pragmatics
Edited by Laurence R. Horn and Gregory Ward

The Handbook of Applied Linguistics
Edited by Alan Davies and Catherine Elder

The Handbook of Speech Perception
Edited by David B. Pisoni and Robert E. Remez

The Handbook of the History of English
Edited by Ans van Kemenade and Bettelou Los

The Handbook of English Linguistics
Edited by Bas Aarts and April McMahon

The Handbook of World Englishes
Edited by Braj B. Kachru; Yamuna Kachru, and Cecil L. Nelson

The Handbook of Educational Linguistics
Edited by Bernard Spolsky and Francis M. Hult

The Handbook of Clinical Linguistics
Edited by Martin J. Ball, Michael R. Perkins, Nicole Müller, and Sara Howard

The Handbook of Pidgin and Creole Studies
Edited by Silvia Kouwenberg and John Victor Singler

The Handbook of Language Teaching
Edited by Michael H. Long and Catherine J. Doughty

The Handbook of Language Contact
Edited by Raymond Hickey

The Handbook of Language and Speech Disorders
Edited by Jack S. Damico, Nicole Müller, and Martin J. Ball

The Handbook of Computational Linguistics and Natural Language Processing
Edited by Alexander Clark, Chris Fox, and Shalom Lappin

The Handbook of Language and Globalization
Edited by Nikolas Coupland

The Handbook of Hispanic Sociolinguistics
Edited by Manuel Díaz-Campos

The Handbook of Language Socialization
Edited by Alessandro Duranti, Elinor Ochs, and Bambi B. Schieffelin

The Handbook of Intercultural Discourse and Communication
Edited by Christina Bratt Paulston, Scott F. Kiesling, and Elizabeth S. Rangel

The Handbook of Historical Sociolinguistics
Edited by Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre

The Handbook of Hispanic Linguistics
Edited by José Ignacio Hualde, Antxon Olarrea, and Erin O'Rourke

The Handbook of Conversation Analysis
Edited by Jack Sidnell and Tanya Stivers

The Handbook of English for Specific Purposes
Edited by Brian Paltridge and Sue Starfield

The Handbook of Spanish Second Language Acquisition
Edited by Kimberly L. Geeslin

The Handbook of Chinese Linguistics
Edited by C.-T. James Huang, Y.-H. Audrey Li, and Andrew Simpson

The Handbook of Language Emergence
Edited by Brian MacWhinney and William O'Grady

The Handbook of Korean Linguistics
Edited by Lucien Brown and Jaehoon Yeon

The Handbook of Speech Production
Edited Melissa A. Redford

The Handbook of Contemporary Semantic Theory, Second Edition
Edited by Shalom Lappin and Chris Fox

The Handbook of Classroom Discourse and Interaction
Edited by Numa Markee

The Handbook of Narrative Analysis
Edited by Anna De Fina and Alexandra Georgakopoulou

The Handbook of English Pronounciation
Edited by Marnie Reed and John M. Levis

The Handbook of Discourse Analysis, Second Edition
Edited by Deborah Tannen, Heidi E. Hamilton, and Deborah Schiffrin

The Handbook of Bilingual and Multilingual Education
Edited by Wayne E. Wright, Sovicheth Boun, and Ofelia García

The Handbook of Portuguese Linguistics
Edited by W. Leo Wetzels, João Costa, and Sergio Menuzzi

The Handbook of Dialectology
Edited by Charles Boberg, John Nerbonne, and Dominic Watt

The Handbook of Linguistics, Second Edition
Edited by Mark Aronoff and Janie Rees-Miller

The Handbook of Linguistics

Second Edition

Edited by

Mark Aronoff and Janie Rees-Miller

WILEY Blackwell

This second edition first published 2017
© 2017 John Wiley & Sons Ltd
Edition history: Blackwell Publishing Ltd (1e, 2001)

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at http://www.wiley.com/go/permissions.

The right of Mark Aronoff and Janie Rees-Miller to be identified as the authors of the editorial material in this work has been asserted in accordance with law.

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty
While the publisher and authors have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data
Names: Aronoff, Mark, editor. | Rees-Miller, Janie, editor.
Title: The handbook of linguistics / Mark Aronoff, Janie Rees-Miller.
Description: Second edition. | Hoboken : Wiley-Blackwell, 2017. | Series: Blackwell handbooks in linguistics | Includes index.
Identifiers: LCCN 2016042791| ISBN 9781405186766 (hardback) | ISBN 9781119072300 (epub)
Subjects: LCSH: Linguistics–Handbooks, manuals, etc. | BISAC: LANGUAGE ARTS & DISCIPLINES / Linguistics / General.
Classification: LCC P121 .H324 2017 | DDC 410–dc23 LC record available at https://lccn.loc.gov/2016042791

Cover image: Mick Tarel/Gettyimages

Set in 9/11pt PalatinoLTStd-Roman by Thomson Digital, Noida, India

10 9 8 7 6 5 4 3 2 1

Contents

List of Contributors ix
Preface to the Second Edition xi
Preface to the First Edition xiii
List of Abbreviations xvii

PART I Starting Points 1
1. Origins of Language
   Andrew Carstairs-McCarthy 3
2. Languages of the World
   Bernard Comrie 21
3. Typology and Universals
   William Croft 39
4. Field Linguistics: Gathering Language Data from Native Speakers
   Pamela Munro 57
5. Writing Systems
   Peter T. Daniels 75

PART II Theoretical Bases 95
6. The History of Linguistics: Approaches to Linguistics
   Lyle Campbell 97
7. Generative Grammar: Rule Systems for Describing Sentence Structure
   Thomas Wasow 119
8. Functional Linguistics: Communicative Functions and Language Structure
   Robert D. Van Valin, Jr. 141

PART III Core Fields 159
9. Linguistic Phonetics: The Sounds of Languages
   John Laver 161

10. Phonology: Sound Structure
Abigail C. Cohn 185
11. Morphology
Andrew Spencer 211
12. The Lexicon
D. A. Cruse 235
13. Syntax
Mark C. Baker 255
14. Formal Semantics
Shalom Lappin 279
15. Historical Linguistics: Language Change Over Time
Brian D. Joseph 299

PART IV Languages and the Mind 321
16. Neurolinguistics
David Caplan 323
17. Psycholinguistics
Kiel Christianson 345
18. Sign Languages
Wendy Sandler and Diane Lillo-Martin 371
19. First Language Acquisition
Brian MacWhinney 397

PART V Languages in Use 415
20. Pragmatics: Language and Communication
Ruth Kempson 417
21. Discourse Analysis
Agnes Weiyun He 445
22. Linguistics and Literature
Nigel Fabb 463
23. Linguistic Anthropology and Ethnolinguistics
Steven P. Black and Elizabeth A. Falconi 479
24. Sociolinguistic Theory: Systematic Study of the Social Uses of Language
J. K. Chambers 505
25. Language Variation: Sociolinguistic Variationist Analysis
Kirk Hazen 519
26. Multilingualism
Suzanne Romaine 541
27. Second Language Acquisition: One Person with Two Languages
Vivian Cook 557

PART VI Applications of Linguistics 583
28. Second Language Pedagogy: Where Theory Meets Practice
Janie Rees-Miller 585
29. Educational Linguistics
James Paul Gee 603

30. Linguistics and Reading
Rebecca Treiman 617
31. Language and Law
Roger W. Shuy 627
32. Translation
Christoph Gutknecht 645
33. Language Planning and Policy
Kathryn D. Stemper and Kendall A. King 655

Index 675

List of Contributors

Mark C. Baker
Rutgers University

Steven P. Black
Georgia State University

Lyle Campbell
University of Hawai'i

David Caplan
Harvard University

Andrew Carstairs-McCarthy
University of Canterbury

J. K. Chambers
University of Toronto

Kiel Christianson
University of Illinois

Abigail C. Cohn
Cornell University

Bernard Comrie
University of California Santa Barbara

Vivian Cook
Newcastle University

William Croft
University of New Mexico

D. A. Cruse
University of Manchester

Peter T. Daniels
Independent Scholar

Nigel Fabb
University of Strathclyde

Elizabeth A. Falconi
Georgia State University

James Paul Gee
Arizona State University

Christoph Gutknecht
University of Hamburg

Kirk Hazen
West Virginia University

Agnes Weiyun He
Stony Brook University

Brian D. Joseph
The Ohio State University

Ruth Kempson
King's College London

Kendall A. King
University of Minnesota

Shalom Lappin
University of Gothenburg

John Laver
Queen Margaret University, Edinburgh

Diane Lillo-Martin
University of Connecticut

Brian MacWhinney
Carnegie Mellon University

Pamela Munro
University of California Los Angeles

Janie Rees-Miller
Marietta College

Suzanne Romaine
University of Oxford

Wendy Sandler
University of Haifa

Roger W. Shuy
Georgetown University

Andrew Spencer
University of Essex

Kathryn D. Stemper
University of Minnesota

Rebecca Treiman
Washington University in Saint Louis

Robert D. Van Valin, Jr.
University at Buffalo

Thomas Wasow
Stanford University

Preface to the Second Edition

The second edition of *The Handbook of Linguistics* has been designed to offer an overview of the field of linguistics in the second decade of the twenty-first century. It has now been over two decades since we began work on the first edition of *The Handbook of Linguistics*, and although our general goals and topics remain much the same as in the first edition, the second edition has been updated to reflect new developments in linguistics since the dawning of the second millennium. New to this edition are chapters devoted to topics not covered in detail in the first edition: psycholinguistics, linguistic anthropology and ethnolinguistics, and second language pedagogy. Other topics that were covered in the first edition have completely new chapters. Specifically, the single chapter devoted to sociolinguistics in the first edition has now been replaced with two new chapters in the second edition, one on sociolinguistic theory and the second on language variation. The topic of language planning has a completely new chapter. Of the remaining chapters, most have been thoroughly revised and updated to incorporate new developments in the field and to refresh any ephemeral examples or references. In a few cases, chapters that have stood the test of time have remained unrevised from the first edition. These are the chapters on languages of the world, history of linguistics, phonetics, the lexicon, and formal semantics.

As in the first edition, the purpose of the *Handbook* is to provide an introduction to the various subfields of linguistics for the educated reader who does not necessarily have a background in linguistics. It is also a useful resource for linguists who may need to teach or to reference subfields outside their specific specializations. In each chapter, we seek to present a broad introduction to the central questions of the subfield and to illustrate how linguists go about answering these questions. Generalizations are supported with enough detail to provide depth, but we have tried to eschew the kind of minutia that would not be meaningful to the general reader or would not stand the test of time.

The order of topics remains much the same as in the first edition. We begin with the starting points for the study of linguistics: the origins of language, the raw material of language study (languages of the world, their typology and universal characteristics, and writing systems), and how language data is gathered from native speakers (field linguistics). The second section of the book considers theoretical bases, beginning with various approaches to the scientific study of language in the history of linguistics. Two chapters are given over to current theoretical perspectives: generative grammar and functional linguistics. We then proceed to the core fields of linguistics, those formal, structural aspects of language that would be covered in almost any general introductory course in linguistics: phonetics, phonology, morphology, the lexicon, syntax,

semantics, and historical linguistics. Chapters in the next section – on neurolinguistics, psycholinguistics, natural sign languages, and first language acquisition – all help to illuminate the relationship between language and the human mind. Our section on language use goes beyond the structural aspects of language to consider how language is used to communicate meaning within various social contexts, in the subfields of pragmatics, discourse analysis, linguistics and literature, linguistic anthropology, and sociolinguistic theory and language variation. The theme of languages in contact is addressed in chapters on multilingualism and second language acquisition. The book ends with chapters concerned with applications of linguistics: second language pedagogy, educational linguistics, linguistics and reading, forensic linguistics, translation, and language planning.

One of the strengths of the first edition of the *Handbook* was that the contributors were internationally recognized scholars in their fields. The same holds true for the second edition: the majority of the authors are senior scholars who contributed to the first edition, while some new chapters have been authored by younger scholars who have emerged as leaders in the field more recently. Our journey to the second edition has not been without detours and bumps in the road, and we are grateful to our contributors and our editors at Wiley Blackwell for their patience and forbearance. We owe a special debt to Agnes He, without whose encouragement, advice, and support this second edition would never have seen the light of day.

Mark Aronoff
Janie Rees-Miller

Preface to the First Edition

For over a century, linguists have been trying to explain linguistics to other people whom they believe should be interested in their subject matter. After all, everyone speaks at least one language and most people have fairly strong views about their own language. The most distinguished scholars in every generation have written general books about language and linguistics targeted at educated laypeople and at scholars in adjacent disciplines, and some of these books have become classics, at least among linguists. The first great American linguist, William Dwight Whitney, published *The Life and Growth of Language: An Outline of Linguistic Science*, in 1875. In the dozen years between 1921 and 1933, the three best known English-speaking linguists in the world (Edward Sapir in 1921, Otto Jespersen in 1922, and Leonard Bloomfield in 1933) all wrote books under the title *Language*. All these books were very successful and continued to be reprinted for many years. In our own time, Noam Chomsky, certainly the most famous of theoretical linguists, has tried to make his ideas on language more accessible in such less technical books as *Language and Mind* (1968) and *Reflections on Language (*1975). And more recently, Steven Pinker's *The Language Instinct* (1995) stayed on the best-seller list for many months.

Despite these efforts, linguistics has not made many inroads into educated public discourse. Although linguists in the last hundred years have uncovered a great deal about human language and how it is acquired and used, the advances and discoveries are still mostly unknown outside a small group of practitioners. Many reasons have been given for this gap between academic and public thinking about language, the most commonly cited being: that people have strong and sometimes erroneous views about language and have little interest in being disabused of their false beliefs; or that people are too close to language to be able to see that it has interesting and complex properties. Whatever the reason, the gap remains and is getting larger the more we learn about language.

The Handbook of Linguistics is a general introductory volume designed to address this gap in knowledge about language. Presupposing no prior knowledge of linguistics, it is intended for people who would like to know what linguistics and its subdisciplines are about. The book was designed to be as nontechnical as possible, while at the same time serving as a repository for what is known about language as we enter the twenty-first century.

If *The Handbook of Linguistics* is to be regarded as authoritative, this will be in large part because of the identity of the authors of the chapters. We have recruited globally recognized leading figures to write each of the chapters. While the culture of academia is such that academic authors find it tremendously difficult to write anything for anyone other than their colleagues,

our central editorial goal has been to avoid this pitfall. Our emphasis on the reader's perspective sets *The Handbook of Linguistics* apart from other similar projects.

The place of the field of linguistics in academia has been debated since its inception. When we look at universities, we may find a linguistics department in either the social sciences or the humanities. When we look at the American government agencies that fund university research, we find that the National Endowment for the Humanities, the National Science Foundation, and the National Institutes of Health all routinely award grants for research in linguistics. So where does linguistics belong? The answer is not in where linguistics is placed administratively, but rather in how linguists think. Here the answer is quite clear: linguists by and large view themselves as scientists and they view their field as a science, the scientific study of language. This has been true since the nineteenth century, when Max Mueller could entitle a book published in 1869 *The Science of Language* and the first chapter of that book "The science of language: one of the physical sciences."

The fact that linguistics is today defined as the scientific study of language carries with it the implicit claim that a science of language is possible, and this alone takes many by surprise. For surely, they say, language, like all human activity, is beyond the scope of true science. Linguists believe that their field is a science because they share the goals of scientific inquiry, which is objective (or more properly intersubjectively accessible) understanding. Once we accept that general view of science as a kind of inquiry, then it should be possible to have a science of anything, so long as it is possible to achieve intersubjectively accessible understanding of that thing. There are, of course, those who deny the possibility of such scientific understanding of anything, but we will not broach that topic here.

We now know that the possibility of scientific understanding depends largely on the complexity and regularity of the object of study. Physics has been so successful because the physical world is, relatively speaking, highly regular and not terribly complex. Human sciences, by contrast, have been much less successful and much slower to produce results, largely because human behavior is so complex and not nearly so regular as is the physical or even the biological world. Language, though, contrasts with other aspects of human behavior precisely in its regularity, what has been called its rule-governed nature. It is precisely this property of language and language-related behavior that has allowed for fairly great progress in our understanding of this delimited area of human behavior. Furthermore, the fact that language is the defining property of humans, that it is shared across all human communities and is manifested in no other species, means that by learning about language we will inevitably also learn about human nature.

Each chapter in this book is designed to describe to the general reader the state of our knowledge at the beginning of the twenty-first century of one aspect of human language. The authors of each chapter have devoted most of their adult lives to the study of this one aspect of language. Together, we believe, these chapters provide a broad yet detailed picture of what is known about language as we move into the new millennium. The chapters are each meant to be freestanding. A reader who is interested in how children acquire language, for example, should be able to turn to Chapter 19 and read it profitably without having to turn first to other chapters for assistance. But the physical nature of a book entails that there be an order of presentation. We begin with general overview chapters that consider the origins of language as species-specific behavior and describe the raw material with which linguists work (languages of the world and writing systems), frame the discipline within its historical context, and look at how linguists acquire new data from previously undescribed languages (field linguistics). The book then turns to the traditional subdisciplines of linguistics. Here we have followed most linguistics books in starting from the bottom, grounding language first in the physical world of sound (phonetics) and moving up through the organization of sound in language (phonology), to the combination of sounds into words (morphology), and the combination of words into sentences (syntax). Meaning (semantics) usually comes next, on the grounds that it operates on words and sentences. These areas are traditionally said to form the core of linguistics, because they deal with the most

formally structured aspects of language. Within the last few decades, however, linguists have come to realize that we cannot understand the most formally structured aspects of language without also understanding the way language is used to convey information (pragmatics) in conversation (discourse) and in literature, and the way language interacts with other aspects of society (sociolinguistics).

Fifty years ago, many of our chapters would have been absent from a book of this sort for the simple but dramatic reason that these fields of inquiry did not exist: language acquisition, multilingualism, sign language, neurolinguistics, computational linguistics, and all of the areas of applied linguistics to which we have devoted separate chapters (the one area of applied linguistics that did exist fifty years ago was language teaching).

The chapters are of a uniform length, approximately 10,000 words each, or about 25 printed pages. This length is substantial enough for a major essay, while being short enough so as not to overwhelm the reader. Applied linguistics is divided into several distinct areas that would be of interest to students and others who want to know what practical applications linguistics has. Because each of the applied linguistics chapters covers a more specialized area, these chapters are somewhat shorter than the rest (approximately 4,000 words each, or about 10 printed pages).

We have tried not to emphasize ideology, but rather to divide things up by empirical criteria having to do with the sorts of phenomena that a given field of inquiry covers. We have thought long and hard about whether some of the major areas, especially syntax and phonology, should be broken down further, with a chapter each on distinct theoretical approaches. Our final decision was not to subdivide by theoretical approaches, based on a belief that the reader's perspective is paramount in books like this: readers of a companion do not want to know what the latest controversy is about or who disagrees with whom or who said what when. Rather, they want to have a reasonable idea of what linguistics or some subarea of linguistics can tell them. The authors have been able to do so without going into the latest controversies, though these controversies may occupy the linguists' everyday lives. The one area to which we have devoted more than one chapter is syntax, but this reflects the dominance of syntactic research in linguistics over the last half century.

We do not see this handbook as an introductory textbook, which would, for example, have questions or exercises at the end of each chapter. There are already enough introductory linguistics texts. We see it rather as an authoritative volume on what linguists know about language at the start of the twenty-first century. Each chapter covers the central questions and goals of a particular subdiscipline, what is generally accepted as known in that area, and how it relates to other areas.

When we embarked on this editorial enterprise, we expected to enjoy the interaction with many of our most distinguished colleagues that the preparation of this book would entail, which is so much easier now in the age of electronic correspondence. What we did not realize was how much we would learn from these colleagues about language and linguistics, simply from reading their work and discussing it with them. We thank all of the authors for this wonderful opportunity and we hope that the readers, too, will share in the same great pleasure.

Mark Aronoff
Janie Rees-Miller

List of Abbreviations

1	first person	GEN	genitive
3	third person	IND	indicative
ABS	absolutive	M	masculine
ACC	accusative	NFUT	nonfuture
ASP	aspect	NOM	nominative
AUX	auxiliary	NPST	nonpast
CLF	classifier	NR	nominalizer
CONN	connective	OBJ	object
COP	copula	P/N	person/number
DAT	dative	PL	plural
EMPH	emphatic	PREP	prepositional
ERG	ergative	PST	past
F	feminine	SG	singular
FUT	future	SUBJ	subject

Part I Starting Points

1 Origins of Language

ANDREW CARSTAIRS-McCARTHY

1 Introduction

Among the inhabitants of some African forests about eight million years ago were ape-like creatures including the common ancestors of chimpanzees and humans. Visualizing these creatures is easy enough; one imagines something resembling a modern gorilla, living substantially in trees and walking on all four limbs when on the ground, and with a vocal communication system limited to perhaps 20 or thirty 30 calls, like a chimpanzee's. But what about our ancestors two million years ago? By that stage they were a separate species from the ancestors of chimpanzees, but were not yet *Homo sapiens*. How did these creatures live, and in particular what sort of language did they have? Visualizing these more recent ancestors is harder. One feels that they must have been more like us, and in particular that their vocal communication system must have been more sophisticated than that of their ancestors six million years before. But how much more sophisticated? Which characteristics of modern human language did this communication system now possess, and which did it still lack?

There is something eerie and yet fascinating about these intermediate ancestors. This fascination underlies innumerable science fiction stories as well as the perennial interest in rumors that such creatures may still exist, in some remote Himalayan valley perhaps, or as descendants of the tiny non*sapiens* humans who may have lived as recently as 15,000 years ago on the island of Flores in Indonesia (Knight 2005; http://en.wikipedia.org/wiki/Homo_floresiensis). To many nonlinguists, therefore, it seems self-evident that research on the linguistic abilities of such intermediate ancestors (that is, research on the origins and evolution of human language) should be a high priority in linguistics. Yet it is not. As a research topic, language evolution is only now beginning to regain respectability, after more than a century of neglect. In the remainder of this section I will say something about the reasons for this neglect before turning in Sections 2 to 5 to the evidence recently brought to bear by anthropologists, geneticists, primatologists, and neurobiologists, who have for decades been more adventurous than linguists in this area. Then in Section 6 I will discuss the kinds of contribution which some linguists also are now beginning to offer.

Many religions provide an account of the origin of language. According to the Judeo-Christian tradition, God gave to Adam in the Garden of Eden dominion over all the animals, and Adam's first exercise of this dominion consisted in naming them. The fact that there are now many languages rather than just one is explained in the story of the Tower of Babel: linguistic diversity is a punishment for human arrogance. So long as that sort of account was generally accepted, the origin of language was not a puzzle. But when secular explanations for natural phenomena

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

began to be sought to supplement or replace religious ones, it was inevitable that a secular explanation would be sought for the origin of language too.

The fact that the origin of language must predate recorded history did not inhibit eighteenth-century thinkers such as Rousseau, Condillac, and Herder, who were confident that simply by applying one's mind to the situation in which languageless humans would find themselves one could arrive at worthwhile conclusions about how language must have arisen. Unfortunately there was no consensus among these conclusions, and in the nineteenth century they came to seem increasingly feeble and speculative by contrast with the far-reaching yet convincing results attainable in historical and comparative linguistics (see Chapter 15). At its foundation in 1866, therefore, the Linguistic Society of Paris banned the presentation of any papers concerning the origin of language. Many linguists still support this ban, in the sense that they believe that any inquiry into the origin of language must inevitably be so speculative as to be worthless.

Since the 1960s, the theory of grammar has come to be dominated by the ideas of Noam Chomsky. For Chomsky, the central question of linguistics is the nature of the innate biological endowment which enables humans to acquire a language so rapidly and efficiently in the first years of life (see Chapter 19). From this viewpoint, it seems natural to regard the origin of language as a matter of evolutionary biology: how did this innate linguistic endowment evolve in humans, and what are its counterparts (if any) in other primates? But Chomsky for a long time discouraged interest in language evolution, and even suggested that language is so different from most other animal characteristics that it may be more a product of physical or chemical processes than of biological ones (1988: 167, 1991: 50). The paradoxical result is that, while Chomskyan linguists endeavored to explain characteristics of individual languages by reference to an innate linguistic endowment (or Universal Grammar), they were generally reluctant to pursue this inquiry one stage further, to the issue of how and why this innate endowment has acquired the particular characteristics that it has. Exceptions (e.g., Newmeyer 1991; Pinker and Bloom 1990; Pinker 1994) were relatively sparse.

In 2002, this situation changed dramatically with the publication of an article jointly written by Chomsky and the animal behavior experts Marc Hauser and Tecumseh Fitch (Hauser, Chomsky, and Fitch 2002). Since then, linguists associated with Chomsky have been willing to discuss language evolution in the context of a general "biolinguistic" exploration of biological bases for the language capacity (see e.g., Jenkins 2004). Their approach is, however, highly controversial (see e.g., Pinker and Jackendoff 2005).

2 Evidence from Anthropology and Archeology

Anthropology is concerned not only with human culture but also with humans as organisms in a biological sense, including their evolutionary development. (On human evolution in general, see, e.g., Stringer and Andrews (2005).) Language is both a cultural phenomenon and also the most salient distinguishing characteristic of modern *Homo sapiens* as a species. The question of how and why humans acquired language therefore interests both cultural and biological anthropologists. So what light can anthropology shed on these questions?

The earliest direct evidence of written language is no more than about 5,000 years old (see Chapter 5). It is therefore much too recent to shed any light on the origin of spoken language, and we must resort to indirect evidence. Unfortunately the available evidence is doubly indirect. The vocal apparatus (tongue, lips, and larynx) of early humans would tell us much if we could examine it directly; but, being soft tissue, it does not survive, and for information about it we have to rely on what we can glean from bones, particularly skulls. Alongside such evidence we have tools and other artefacts, as well as traces of human habitation such as discarded animal bones; but, again, what is available to us is skewed by the fact that stone survives better than bone and much better than materials such as wood or hide. In view of this, the only relatively

firm dates which anthropology can provide are two terminuses, one after which we can be sure that language in its fully modern form did exist and one before which we can be sure that it did not. For the long period in between, the anthropological evidence is tantalizing but frustratingly equivocal; there are no uncontroversial counterparts in the fossil record for specific stages in linguistic evolution.

We can be reasonably confident that modern-style spoken language evolved only once. This is not logically necessary. It is conceivable that something with the communicative and cognitive functions of language, and using speech as its medium, could have evolved independently more than once, just as the eye has evolved independently more than once in the animal kingdom. However, if that had happened we would expect to find evidence of it today. We would expect to find two or more different kinds of language, differing in structure in such a way that people biologically disposed to learn one kind would never be able to acquire another kind natively. These would be differences as fundamental as those between the eyes of octopuses, mammals, and insects. Yet no such evidence exists. For all their diversity, all existing languages display certain fundamental common properties of grammar, meaning, and sound. For this reason Chomsky feel justified in claiming that, to a visitor from another planet, it might seem that there really is only one human language. Moreover, a child who is removed from her parents' speech community at a young age can acquire natively any language whatsoever, irrespective of what her parents speak. There is no evidence that any child is born with a biological bias in favor of one language or type of language. This means that language of a fully modern kind must have evolved before any contemporary human group became geographically separated from the rest of the human race (separated, that is, until the invention of modern means of transport). The first such clearcut separation seems to have occurred with the earliest settlement of Australia by *Homo sapiens*. Archeological evidence suggests that that event took place at least 40,000 years and perhaps as long as 60,000 or more years ago. We can therefore take this as a firm *terminus ante quem* for the evolution of a form of language which is fully modern in a biological sense.

As for a *terminus post quem*, it is clear that spoken language with more or less modern articulatory and acoustic characteristics presupposes something like a modern vocal tract. But how are we to interpret "more or less" and "something like"? One thing is clear: the acoustic properties of many human speech sounds, particularly vowels, depend on the characteristically human L-shaped vocal tract, with an oral cavity at right angles to the pharynx (see Chapter 9) and with the larynx relatively low in the neck. This shape is characteristically human because in nearly all other mammals, and even in human babies during the first few months of life, the larynx is high enough for the epiglottis to engage with the soft palate so as to form a self-contained airway from the nose to the lungs, smoothly curved rather than L-shaped, and quite separate from the tube which leads from the mouth to the stomach. Having these two distinct tubes enables nearly all other mammals, as well as newborn human babies, to breathe while swallowing. The adult human pharynx, on the other hand, through which both air and food must pass, contributes importantly to the acoustic characteristics of speech sounds. So when did this L-shaped vocal tract develop?

Lieberman (1984; cf. Lieberman and Crelin 1971) has claimed that even in Neanderthals, who did not become extinct until about 35,000 years ago, the larynx was positioned so high in the neck as to prevent the production of the full modern range of vowel sounds. He suggests that this linguistic disadvantage may have been a factor in the Neanderthals' demise. But his argument rests on an interpretation of fossil cranial anatomy which has generally been rejected by anthropologists (Trinkaus and Shipman 1993; Aiello and Dean 1990). An alternative view is that the L-shaped vocal tract is a byproduct of bipedalism, which favored a reorientation of the head in relation to the spine and hence a shortening of the base of the skull, so that the larynx had to be squeezed downward into the neck (DuBrul 1958; Aiello 1996b). The question then arises: when did our ancestors become bipedal? The general consensus among anthropologists is: very early. Evidence includes fossil footprints at Laetoli in Tanzania, from about 3.5 million years ago, and the skeleton of *Australopithecus afarensis* nicknamed "Lucy," dating from over three million years

ago. So, if bipedalism was an important factor contributing to the lowering of the larynx, the L-shaped vocal tract probably emerged relatively early too.

This conflicts with an opinion widespread among language origin researchers, namely that the lowering of the larynx (with its concomitant increased risk of choking) was a consequence of the evolution of more sophisticated language, not a precursor of it. This "brain-first" view was inevitably popular so long as Piltdown Man, with its human-like skull and ape-like jaw, was believed to be genuine. More recent evidence, showing how small australopithecine and early human skulls were, seems to count against the "brain-first" view. On the other hand, in yet more recent work, Fitch (2002) and others have shown that the lowered larynx is not so unusual among nonhuman mammals as was once thought. This in turn suggests that, whatever the reasons are why language is uniquely human, the vocal apparatus may not after all be centrally important.

Mention of skulls raises the possibility of drawing conclusions about language from hominid brains. (I use the term "hominid" to mean "(belonging to) a creature of the genus *Australopithecus* or the genus *Homo*.") Brain size tells us nothing specific. But what of brain structure? If it could be shown that an area of the modern human brain uniquely associated with language was present in the brains of hominids at a particular date, it would seem reasonable to conclude that those hominids possessed language. But this line of reasoning encounters three problems. Firstly, since brains themselves do not fossilize, determining their structure depends on the interpretation of ridges and grooves on the inside of skulls, or rather of their counterparts on "endocasts" made from skulls. The region generally regarded as most closely associated with grammar and with speech articulation in modern humans is Broca's area; but identifying an area corresponding to Broca's area in hominid fossils has turned out to be highly controversial (Falk 1992). Secondly, no area of the human brain, even Broca's area, seems to be associated with language and nothing else. Thirdly, Broca's area seems to have little or nothing to do with vocalization in monkeys, so even if it can be established that a counterpart of Broca's area exists in a certain hominid, its function in that hominid may not be linguistic. We will discuss Broca's area again in Section 5. For the time being, though, the details of "brain-language coevolution," as Deacon (1997) calls it, remain frustratingly indeterminate.

Some scholars have connected language with the evolution of "handedness," which is much more strongly developed in humans than in other animals (Bradshaw and Rogers 1992; Corballis 2002). In most people the right hand is the dominant hand, controlled from the left side of the brain where the language areas are usually located. It is tempting to see this shared location as more than mere coincidence. If so, linguistic conclusions might perhaps be drawn from ingenious tests that have been carried out on fossil stone tools, to determine whether the people who made them were or were not predominantly right-handed. However, the correlation between language and handedness is far from strong: left-handedness neither entails nor is entailed by right-brain dominance for language. Also, even if evidence of a strong preponderance of right-handers in some group of hominids is taken as firm evidence of linguistic capacity, it furnishes no details about the nature of that linguistic capacity.

Let us turn from biology to culture. Common sense would suggest that a relatively sudden jump in the complexity of human linguistic behavior, if it occurred, should leave immediate traces in the archeological record in the shape of a sudden jump in the complexity of preserved artefacts (tools, ornaments, and artwork). So does any such jump in complexity occur, and if so, when? There is indeed a big increase in the variety and quality of tools found in Europe and Africa around 40,000 years ago, followed by the famous cave paintings of Lascaux and elsewhere from about 30,000 years ago. But this is inconveniently late as a date for the emergence of fully modern language, in that it is contemporary with or even more recent than the latest plausible date for the settlement of Australia. That has not discouraged some scholars from using this kind of evidence to argue that language evolved "late"; but on examination it generally turns out that what these scholars mean by "language" is not what linguists mean by it, but rather the self-conscious use of symbols (Noble and Davidson 1996). Moreover, there is scattered but intriguing

evidence of "cultural" behavior thousands of years earlier, such as burial pits, incised bones, and the use of red ocher pigment for body decoration. The linguistic implications of this for language are unclear, but it may be significant that some of the dates involved are not far removed from a milestone indicated by genetic evidence, to which we now turn.

3 Genetic Evidence

Within the last three decades, molecular genetics has opened up entirely new techniques for assessing the relationship of humans to each other and to other primates. (Genetic evidence tells us that we are separated by only about five million years from the ancestor which we share with the chimpanzees.) Since the 1950s it has been known that the information which differentiates an individual genetically from all other individuals (except a possible identical twin) is carried by DNA (deoxyribonucleic acid) in chromosomes located in every cell in the body. Geneticists can now compare individuals and groups in terms of how much of their DNA is shared. Moreover, they can do this not only with respect to the DNA in the cell's nucleus, which is inherited from both parents, but also with respect to the DNA in the cell's mitochondria – some of the so-called "organelles" which the cell contains in addition to its nucleus. What is important about mitochondrial DNA is that it is inherited from the mother alone. It follows that the only reason that there can be for any difference between two people's mitochondrial DNA is inaccurate inheritance due to mutation; for, without this inaccuracy, both of them would have exactly the same mitochondrial DNA as their most recent shared ancestor in the female line. So, assuming that mutation in DNA occurs at a constant rate, the extent of difference between two people's DNA is an indication of the number of generations which separate them from the most recent woman from whom both are descended through her daughters, her daughters' daughters, and so on.

Cann, Stoneking, and Wilson (1987) used this technique to try to locate in time and space the most recent woman from whom all living humans are descended in the female line. With the help of elaborate statistical techniques, they argued that this woman lived roughly 200,000 years ago in Africa, hence the nickname "African Eve." Both the African location and the date corresponded quite closely to the "out-of-Africa" scenario for early *Homo sapiens* proposed on independent grounds by some archeologists, so the two theories provided mutual support. The nickname "Eve" is convenient but unfortunate, because it suggests that, apart from Eve's male partner or partners, none of her contemporaries has any descendants alive today. That is a fallacy; all one can say is that anyone alive today who is descended from a female contemporary of Eve must be linked to that woman through at least one male ancestor. However, the argument of Cann and her colleagues does suggest that there was a population bottleneck relatively recently in human prehistory, such that most of the humans alive around 200,000 years ago, scattered over large areas of Africa, Europe, and Asia, have indeed left no surviving descendants. Why should this be?

Many scholars have been tempted to suggest that what was special about Eve's community – the characteristic which enabled their descendants to outperform other humans and which discouraged interbreeding with them – must have been superior linguistic abilities, presumably newly acquired. This is only a guess, however. Cann herself has more recently mentioned one of many alternative possibilities: infectious disease (Cann, Rickards, and Koji-Lum 1994). But the possible link with language evolution has been popularized by Cavalli-Sforza (1995) and by Ruhlen (1994), whose supposed reconstructions of Proto-World vocabulary might, if genuine, be roughly contemporary with Eve. An equivocation on "mother tongue" underlies this view, however. Even supposing it were possible to reconstruct the most recent language from which all contemporary languages are descended, it would be a remarkable coincidence if that ancestral language (the "mother tongue" in a historical linguistic sense) were also the first linguistic variety with fully modern characteristics (the "mother tongue" in a biological sense). So, once again, we are faced with evidence which, though tantalizing, does not point to any firm conclusion.

Another recent discovery has been evidence for a relationship between inherited language impairment and a specific gene: the so-called FOXP2 gene (Lai et al. 2001). So is FOXP2 "the language gene"? No serious scholar now makes that claim. For one thing, it is found also in the DNA of many other species. Its role in language is at best indirect, and many other genes are relevant to the normal maturation of language in humans. However, it seems possible that a specifically human change in the FOXP2 gene around 200,000 years ago may turn out to be an important piece in the jigsaw, because that is a plausible date for the beginning of the territorial expansion of *Homo sapiens* beyond southern Africa.

4 Primatological Evidence

No living primate apart from man is equipped to speak. However, three areas of current research on primates may shed light on language evolution. These involve primate vocal call systems, primate cognitive abilities (particularly their knowledge of social relationships), and the results of experiments involving teaching sign language and artificial signaling systems to apes.

4.1 Vocal call systems

Until a couple of decades ago, it was generally thought that the calls uttered by all animals, including monkeys and apes, were exclusively reflections of physical or emotional states such as pain, fear, hunger or lust. In this respect, the portion of the human vocal repertoire which primate call systems seemed to resemble most closely was the portion consisting of involuntary sounds such as cries of pain, laughter or sobbing. No linguists have been reluctant to contemplate an evolutionary link between these cries and primate vocalizations. But primate "vocabularies" were thought to lack a central element of human vocabularies: referential calls identifiable with specific objects or classes of objects in the external world. Given that assumption, it was easy to dismiss animal call systems as irrelevant to human language. However, students of animal behavior were becoming increasingly uncomfortable with this assumption, and Cheney and Seyfarth (1990) developed a particularly elegant and convincing way of testing it systematically. (On animal communication generally, see Hauser 1996; on the calls of chimpanzees in the wild, see Goodall 1986.)

In the 1970s and 1980s, Cheney and Seyfarth spent years investigating the behavior of vervet monkeys in their native habitat, the Amboseli National Park of Kenya. These small monkeys utter distinct warning calls for different types of predator, notably leopards, snakes, and eagles, for which different types of evasive action are appropriate: they run up trees to escape leopards, peer at the ground around them to avoid snakes, and hide in bushes to evade eagles. This kind of apparent referentiality had been noticed before, not just among vervets; but such awareness had not shaken the general conviction among both zoologists and linguists that animal cries were basically emotional or affective in content rather than referential. In crude terms, a vervet's eagle call would be interpreted as linked not to something in the outside world ("There's an eagle!") but rather to its internal state ("I am experiencing eagle-fear!" or "I feel an urge to hide in bushes!"). To be sure, if one vervet uttered the eagle call, others might take evasive action too; but this could only be because these others saw the eagle for themselves and hence experienced the same emotion (it was thought).

Cheney and Seyfarth showed this interpretation to be incorrect by way of a crucial experiment. They made recordings of predator warning calls and played them back from hidden loudspeakers in the absence of the relevant predators. If the traditional interpretation of the warning calls was correct, the vervets would be predicted to take no evasive action in response to these bogus calls. They might look for around for the relevant predator but, failing to see one, they would not experience the relevant fear reaction and so would do nothing. However, what Cheney

and Seyfarth found was that the vervets reacted to the bogus calls just as if they were genuine, by taking the appropriate evasive action. The call itself was the trigger to act, not the emotion or physical state engendered by the sight of a predator. Warning calls therefore really do contain referential information about the environment, on which vervets can act appropriately. To this admittedly limited extent, therefore, they resemble words of a human language.

A second respect in which human language differs from animal cries, it used to be thought, is that only human language can be unreliable. If an animal cry is an automatic response to an emotional or physical stimulus, its reliability is in some sense guaranteed. Humans, on the other hand, can tell lies or make mistakes. But Cheney and Seyfarth showed that in this respect too the gap between vervet monkeys' calls and human language is less than was once thought. Vervets' use of their warning calls is not entirely innately determined; for example, young vervets will sometimes utter the eagle call even when they have seen something in the sky which is not an eagle or even a bird at all, such as a falling leaf. And adult vervets react differently to young vervets' calls too. Instead of taking immediate evasive action, as they would if they had heard an adult call, they first check for themselves whether the relevant predator is present and, if not, ignore the call. It seems to be through observing when its calls are acted upon and when they are ignored that a young vervet refines its innate repertoire of vocal reactions into accurate warnings deployed according to the conventions of the adult community.

These observations show that, for vervets, calls have a content which is independent of their own physical or emotional state. Cheney and Seyfarth were also able to show that, in judging the reliability of a call that it hears, a vervet goes beyond merely identifying the caller. It is clear that vervets can distinguish individual "voices," because when a young vervet utters a cry of distress the adults in earshot will look toward that individual's mother, as if expecting her to respond. Cheney and Seyfarth compared reactions to recordings of different voices uttering a variety of calls. In the absence of a genuine eagle danger, hearers will become habituated to and hence ignore recorded eagle alarms in the voice of vervet A, but will still react to alarms in the voice of vervet B. But, even when so habituated to vervet A, they will not ignore a recording of vervet A uttering a call of a different kind (say one of the repertoire of calls relating to individual or group interactions). Vervets can evidently distinguish, in respect of another vervet, those topics on which it is a reliable witness from those on which it is unreliable.

To be sure, the vervet call system has no grammatical organization remotely resembling that of human language, and the same is true of all other primate call systems. Nevertheless, the observations of Cheney, Seyfarth, and others tend to show that the differences between primate call systems and human language are not so great as was once thought, and hence weaken the case for denying any evolutionary connection between them.

4.2 Cognitive abilities

Longterm observations of primate groups in the wild, such as those of Goodall and Cheney and Seyfarth mentioned in Section 4.1, show that primates know many more details about themselves, their conspecifics, and their environment than was previously suspected. In particular, they can distinguish kin from nonkin, and by remembering who has done what to whom they can distinguish allies from enemies. This is relevant to language inasmuch as a fundamental characteristic of language is the ability to represent grammatically the roles of participants in a situation (Bickerton 1990, 2002). For example, the sentence *John gave Mary a banana* represents a situation in which John is the agent, Mary is the goal, and the banana is the patient or "theme" in relation to an act of giving. In the terminology of semantics, such a set of relationships between participants in a situation is called a "thematic structure" or "argument structure" (see Chapter 7). Higher primates do not produce sentences, but they certainly have mental representations of thematic structures of the kind which underlie sentences. To that extent they have evolved to a stage of cognitive readiness for language.

One of the rubicons which have been claimed to separate humans from other animals is that, whereas other animals may possess "procedural" knowledge ("knowledge-how"), only humans have access to "propositional" knowledge ("knowledge-that"). (In a similar vein, Donald (1991) distinguishes between "episodic," "mimetic," and "mythic" culture, among which only "episodic" culture is available to nonhumans.) If this is correct, it is tempting to see propositional knowledge as a prerequisite for language. In assessing whether this is correct, however, one immediately encounters a risk of circularity. If "propositional knowledge" means simply "knowledge of a kind which can only be represented in sentence form," then it is not surprising that propositional knowledge should be restricted to sentence-users, that is, to humans; but then to say that animals lack it is to say no more than that animals lack language. On the other hand, if "propositional knowledge" is defined so as to make it logically independent of language, such as in terms of thematic structure, it is by no means so clear that this rubicon exists.

At least two considerations support the idea that primates have access to "knowledge-that." One is the extent to which, in the admittedly artificial conditions of the laboratory, chimpanzees can acquire and display awareness of abstract concepts such as "same" and "different" and apply them by reference to a range of criteria such as color and size (Premack 1976). Even more striking is the ability of macaque monkeys to memorize long sequences of symbols and touch them rapidly on a screen in the right order, even when the arrangement of them changes each time, in order to get food rewards (Terrace 2002). (Imagine if, every time you withdrew money from an automatic teller machine, you found the numbers in a different configuration!) The other consideration is the fact that primates can apparently indulge in deception, or display what has been called "Machiavellian intelligence" (Byrne and Whiten 1988; Sommer 1992). In interpreting "Machiavellian" behavior it is of course necessary to guard against overenthusiastic ascription of human personality traits to animals. Nevertheless, this behavior suggests that primates are capable of conceiving of situations which do not exist, that is, to think in an abstract "propositional" fashion, and hence reinforces the worthwhileness of looking for precursors of language in other species.

Social relationships among primates are both more complex and less stereotyped than among other mammals, and it has been suggested that social factors may outweigh communicative ones in fostering language evolution. Dunbar (1996) and others have drawn attention to the relationship between group size, brain size, and social grooming in various primate species. Grooming is important in fostering group cohesion; on the other hand, time devoted to grooming increases exponentially as group sizes increase, thereby reducing the time available for other essential tasks such as food gathering. Dunbar suggests that language provided a way out of this dilemma: it is a form of vocal grooming, with the advantage that by means of language one can groom many other individuals at once. Traces of this original function can be observed in the extent to which, even today, language is used for gossip and for cementing social relationships rather than for the more abstract representational and information-conveying purposes which tend to interest grammatical theorists and philosophers.

4.3 Sign language experiments

Apes do not have vocal tracts suited for speech, but their arms and hands are physically quite capable of forming the signs of Deaf languages such as American Sign Language (ASL). In the 1970s great excitement was generated by experiments which purported to show that chimpanzees could learn ASL, so that language could no longer be regarded as a uniquely human attribute (Terrace 1979; Gardner, Gardner, and Van Cantfort 1989). Linguists in general denied hotly that the sign sequences produced by chimpanzees such as Washoe and Nim could be regarded as genuine syntactic combinations or complex words, pointing to the fact that the chimpanzees' sign sequences never reached the variety and complexity of those of fluent human ASL signers.

The chimpanzees' supporters, on the other hand, argued that the kinds of sign combination which chimpanzees produced were quite similar to the word combinations which human babies produce at the "two-word" or "telegraphic" stage of language acquisition, so that, if what the chimpanzees did was not a manifestation of language, one could not call babies' "telegraphic" speech a manifestation of language either. (We will return to this implication in Section 6.) In the present context the issue is not whether the chimpanzees' and other apes' signing behavior can properly be called linguistic (which risks becoming a sterile issue of terminology), but whether this behavior sheds any light on language evolution.

One effect of the ape language experiments was to give new life to the old idea that language in humans may have originated in gesture, and only later been transferred to the vocal channel (Armstrong, Stokoe, and Wilcox 1995). Just as apes can sign without a human vocal tract, so could our australopithecine ancestors have communicated by sign before their vocal tracts had become capable of modern-style speech, perhaps. One of the attractions of this proposal has always been that it seems to provide a solution to the problem of how humans originally learned to handle the arbitrary relationship between words and meanings. The apparent solution lies in the fact that many signs in ASL and other sign languages are motivated ("iconic") rather than arbitrary ("symbolic"), that is, they resemble or recall in some way their referents in the outside world, while many other signs were once more clearly motivated than they are now. The proportion of sign language vocabularies which is iconic is far greater than the proportion of iconic (onomatopeic) words in spoken language vocabularies. These motivated manual signs could have constituted a scaffolding, so to speak, to assist the more difficult task of mastering arbitrary signs, whether manual or vocal. But the attraction of this reasoning disappears as soon as one recalls that vervet monkeys' call vocabulary is just as symbolic as most words of human language. Vervets' eagle, leopard, and snake calls do not in any way resemble or sound like eagles, leopards or snakes. So, even if one regards the use of symbolic signs as a communicative rubicon, it is a rubicon which has been crossed by any nonhuman species with a clearly referential call vocabulary, and was almost certainly crossed by our primate ancestors long before the appearance of hominids.

More relevant to language evolution, perhaps, is what can be gleaned from observation of the bonobo (or pygmy chimpanzee) Kanzi (Savage-Rumbaugh et al. 1993; Savage-Rumbaugh and Lewin 1994). Savage-Rumbaugh set out to train Kanzi's mother in both sign language and the use of a keyboard of arbitrary wordsigns or "lexigrams," while the infant Kanzi was left to play and watch what was going on unmolested. The mother turned out to be an unpromising pupil. Kanzi, on the other hand, developed spontaneously a form of communication involving both manual signs and lexigrams, and also showed a surprising ability to understand spoken English – a somewhat more accurate understanding, in fact, than the two-year-old daughter of one of Savage-Rumbaugh's colleagues, at least within a deliberately limited range of syntactic constructions.

Savage-Rumbaugh argues that Kanzi shows evidence of rule-governed use of signs and lexigrams, and one may if one wishes call this set of rules a syntax. But it seems overhasty to conclude, as Savage-Rumbaugh does, that the difference between Kanzi's syntax and that of human languages is only in degree of complexity, not in kind. Of the two rules which Kanzi has invented rather than merely copied from human sign use, one ("lexigram precedes gesture") clearly has no human counterpart, while the other ("action precedes action, sign order corresponding to order of performance," as in *chase hide* or *tickle bite*) is interpretable as purely semantic or pragmatic rather than syntactic. Moreover, Kanzi's "utterances" are nearly all too short to permit clearcut identification of human-language-like phrases or clauses. A more conservative conclusion would be that Kanzi may indeed have invented a kind of rudimentary syntax, but it cannot be straightforwardly equated with the kind of syntax that human languages have. A task for the language evolution researcher, then, is to account for the differences between what the bonobo does and what humans do.

5 Neurobiological Evidence

To investigate systematically the relationship between language and the brain, one would need to carry out surgical experiments of an ethically unthinkable kind. Our knowledge has therefore to be gleaned in a relatively haphazard fashion, from the linguistic behavior of people suffering from brain damage due to accident or cerebral hemorrhage. This is less than ideal, because the extent of the damage is of course not subject to any experimental control and is determinable only indirectly, through methods such as magnetic resonance imaging (MRI), which is like an X-ray but much more detailed, and positron emission tomography (PET), which measures minute changes in bloodflow. With the patient's consent, it is also possible to test the linguistic effect of stimulating areas of brain tissue directly in the course of surgery for purposes such as the control of epilepsy (Calvin and Ojemann 1994). Not surprisingly, the literature on such research, though extensive, is somewhat confusing. However, it does suggest answers (though by no means conclusive ones) to two broad questions relevant to language evolution. The first question concerns the relative priority of the vocal and gestural channels for speech. The second concerns the extent to which syntax is an outgrowth of a general increase in the sophistication of hominids' mental representation of the world, including social relationships, and the extent to which it is an outgrowth of some more specialized development, such as better toolmaking, more accurate stone-throwing, or more fluent vocalization.

Before we consider these broad questions, it is worth emphasizing that the relationship between particular functions and particular brain locations is not clearcut and unchanging, either in the individual or in the species. Exercising one finger can increase the area of brain cortex devoted to controlling it, and in many blind people the cortex areas for finger control are larger than average. This functional plasticity is particularly evident in early infancy, so that a young child who suffers massive damage to the left brain hemisphere (where the control of language is generally located) may nevertheless acquire a considerable linguistic capacity, controlled from the right hemisphere. Indeed, without such plasticity and scope for functional overlapping it is hard to see how language could have evolved at all, because it must have involved a new role for parts of the brain which originally served other functions.

The brain region which seems most clearly implicated in regulating grammar is Broca's area, in the frontal lobe of the left hemisphere. In view of the scope for overlap in functions, it seems reasonable to predict that, if language was originally gestural, Broca's area would be relatively close to that part of the brain which controls movement of the hands; but it is not. Control of bodily movements resides on the so-called motor strip, just in front of the central sulcus or Rolandic fissure which separates the frontal lobe from the parietal lobe. Broca's area is indeed close to the motor strip; but it is closest to that part of the strip which controls not the hands but rather the tongue, jaw and lips. Moreover, a similarly located Broca's area seems to be just as relevant to the grammar of sign language, even among people deaf from birth, as it is to the grammar of spoken language (Poizner, Klima, and Bellugi 1987).

Conceivably the region for grammatical control could have migrated, so to speak, if the predominant channel for language switched from gesture to speech. However, since the present location of Broca's area does not prevent it from playing a role in sign language, a hypothetical language area located close to the manual section of the motor strip could presumably have retained its original control over grammar even while the vocal apparatus took over from the hands. So it seems more likely that the linguistic function exercised by Broca's area has not migrated, and its present brain location reflects the fact that human language has always been predominantly vocal.

Damage to Broca's area affects grammar and speech production more than it affects vocabulary. Broca's aphasics can generally produce appropriate nouns, adjectives, and verbs for what

they are trying to say; it is the task of stringing them together in well-formed sentences with appropriate grammatical words (determiners, auxiliaries, and so on) which causes them trouble. A complementary kind of aphasia, involving fluent grammar but inappropriate or nonsensical vocabulary, is associated with damage elsewhere in the left hemisphere, in a region of the temporal lobe and part of the parietal lobe known as Wernicke's area. In Wernicke's aphasics the grammatical equipment to talk about the world may be intact, but access to the concepts for organizing their experience of the world (insofar as one can equate concepts with items of vocabulary) is disrupted. Wernicke's aphasia is therefore problematic for any suggestion that conceptual relationships such as thematic structures (mentioned earlier) were not merely a necessary condition for the evolution of syntax but rather the main trigger for it. On the basis of that suggestion, one would expect lexical and grammatical disruption regularly to go hand in hand, rather than to occur independently. So, in answer to our second question, the characteristics of Wernicke's aphasia suggest that, for syntax to evolve as it has, something more specialized than just general conceptual sophistication was necessary.

Various suggestions have been made concerning this more specialized ingredient. Some scholars have appealed to the hierarchical organization of relatively complex behaviors involving tools (e.g., Greenfield 1991). Calvin (1993) has pointed out the neurobiological advances necessary for muscular control in accurate throwing, and has suggested that the relevant neural structures may have been coopted for rapid, effortless syntactic organization of words in speech. But the most influential of such suggestions in recent years rests on the discovery of "mirror neurons" (Arbib 2006). In monkeys, what corresponds to Broca's area has a role in the control of activities involving the hand and arm. But it is not just a monkey engaged in such activities whose Broca's area lights up (so to speak). Corresponding neurons light up in the Broca's area of any watching monkey, "mirroring" what happens in the brain of the monkey performing the action. Could it be, then, that this neural mirroring led to imitative gesturing, which could then have led to human language (still closely associated with Broca's area) via a gestural route? One difficulty with this idea is that, if gestural mimicry played an important role in the evolution of language, one would expect direct mimicry (vocal if not gestural) still to play an important role in early language acquisition. Yet, notoriously, young children are unwilling mimics in that they resist formal correction of their grammatical errors. It is as if their brains are equipped to get things right in their own good time, and not before. So mirror neurons constitute yet another jigsaw puzzle piece whose proper position remains uncertain.

6 Linguistic Evidence

It may seem paradoxical that the section on linguistic evidence for the origins of language has been left until last. However, as explained in Section 1, linguists have been relative latecomers to this field. Their contributions can be divided into those focusing on the relationship between language and "protolanguage" and a more recent and disparate group focusing on the evolutionary rationale for particular aspects of modern grammatical organization.

6.1 Protolanguage and "true" language

Students of language contact distinguish between pidgins, which are used as second languages in situations of regular contact between people with mutually unintelligible mother tongues, and creoles, which arise when children acquire pidgins natively. The creolization process involves faster spoken delivery and the rapid appearance of new grammatical features which may be expressed unsystematically or not at all in the parent pidgin. Study of creole formation, especially among children of workers on Hawaiian sugar plantations, led Bickerton (1981) to the

controversial proposal that, in environments where creoles originate, the universal human linguistic "bioprogram" reveals its characteristics most plainly, because the local speech community lacks entrenched grammatical habits which might interfere with it.

Since proposing the bioprogram hypothesis, Bickerton has turned his attention to how the bioprogram may have evolved, and to what sort of linguistic capacity may have preceded it (1990; 1995). He has suggested that what preceded it is still present and in use among humans in certain situations: in "touristese," in the speech of people who are intoxicated or suffering from some kinds of brain damage, and especially in the "two-word" or "telegraphic" stage of infant speech already mentioned in Section 4. This kind of language lacks any systematic grammar, so to understand it one must rely heavily on semantic and pragmatic cues. In particular, it lacks any systematic encoding of thematic structure of the kind which, in "true" language, allows us to distinguish reliably between agents, patients, beneficiaries, instruments, and so on (see Section 4). In the English sentence *John killed a crocodile*, the identity of the agent and the patient is reliably indicated by word order, while in the Latin sentence *Johannes crocodilum interfecit* it is the endings *-s* and *-m* which serve this purpose; however, on hearing an English-based protolanguage utterance such as *John crocodile kill* one cannot know whether to mourn or rejoice without the help of contextual or background knowledge.

An alternative to Bickerton's version of the bioprogram hypothesis is presented by Wray (1998). She suggests that protolanguage utterances were "holistic" rather than wordlike; they encoded whole messages (e.g., "The baby wants milk") rather than individual proto-nouns or proto-verbs (e.g., "baby," "milk," "want"). In favor of this view is that it renders protolanguage utterances more similar in character to primate vocalizations, from which they perhaps developed. It is also true that, even in contemporary language, many apparently complex expressions are stored and retrieved "holistically," e.g., *That's the way the cookie crumbles*, or *Do me a favor!* with the meaning "Don't expect me to believe that!" On the other hand, there are difficulties with Wray's account of how these holistic utterances might later have come to be analyzed as containing meaningful parts (Tallerman 2007).

One striking fact about hominid evolution is that increase in brain size was not steady. Rather, there was a first burst of brain expansion between 2 and 1.5 million years ago, as *Homo habilis* and *Homo erectus* came to replace the earlier australopithecines, followed by a second burst within about the last 300,000 years, as *Homo sapiens* came to replace *Homo erectus* (Aiello 1996a). Various factors, such as diet and group size, have been invoked to explain this. Bickerton's approach to the problem is to ask why *Homo erectus*, though capable of quite sophisticated toolmaking, failed to make any significant technological or cultural advance for over a million years. His answer is that *Homo erectus* was endowed not with "true" language but only with protolanguage. Those hominids were at least as aware of social relationships as present-day apes are, and could represent thematic structures (who did what to whom) mentally; but they had no reliable linguistic tool for talking about these relationships or expressing these mentally represented structures. Linguistically, they were trapped throughout their lives at the two-word stage of the modern toddler.

Bickerton thus provides an intriguing, though speculative, answer to the question of what held *Homo erectus* back for so long. But how did humans ever get beyond protolanguage? Bickerton's answer is that new neural connections in the brain allowed speech to be hooked up to thematic structure. This would have yielded a sudden and dramatic improvement in the reliability and versatility of language, and hence set the stage for the rapid advances of the last quarter of a million years. What is less clear is why the neural hookup should have occurred when it did, rather than earlier or later. Bickerton's scenario also supplies no particular reason why grammar should be more closely associated in the brain with control of the vocal tract than with the organization of vocabulary. But his proposals certainly suggest one way of reconciling the Chomskyan view of modern human language as qualitatively unique with the need to accommodate it somehow in an account of human evolution.

6.2 Actual grammar versus conceivable grammars

Is the sort of grammar that languages have the only kind that they could conceivably have, or does grammar-as-it-is represent only one of many directions which linguistic evolution might have taken? This is an ambitious question, and there is no guarantee that it can be answered; however, it is the sort of question which only linguists, among the various contributors to language evolution studies, are equipped to tackle.

If one says that the characteristics of grammar-as-it-is are inevitable, one is saying in effect that grammar is as it is for the same sort of reason that the seed compartments in a pine cone, or the individual florets in a sunflower, are distributed in a pattern corresponding to the Fibonacci series, in which each number is the sum of the previous two (0 + 1 + 1 + 2 + 3 + 5 + 8 + 13 + 21 + …). Reasons of this kind reflect not biology but physics or mathematics. That is the kind of possibility that Chomsky has in mind when he says that the ability to learn grammars "may well have arisen as a concomitant of structural properties of the brain that developed for other reasons" (quoted by Pinker 1994: 362). A more recent suggestion of Chomsky's is that there may be nothing that is both peculiar to language and peculiar to humans except recursion (the capacity to embed one structure, such as a sentence, into another of the same kind) (Hauser, Chomsky, and Fitch 2002). The mental capacity for recursion is probably a peculiarly human development and may well have arisen suddenly. This has intriguing implications for what humans do and do not share with other animals.

A number of researchers are now using computer simulation to explore what happens when a signaling system with certain initial characteristics is set up to be adaptable so as to fit better the needs of the system's "users" (Batali 1998; Berwick et al. 1998; Steels 1997; Chapters 15–17 in Christiansen and Kirby 2003). If common trends emerge from these experiments, and if these trends correspond to identifiable aspects of grammar and vocabulary, that may indicate that the aspects in question were bound to evolve as they have, irrespective of the fact that it is in the language of humans that they appear rather than in a "language" used by dolphins or Martians. Any firm findings in this line lie in the future, however.

What of aspects of grammar which are not inevitable in this sense? A central issue is whether or not all aspects of grammar are well-engineered responses to selection pressures to which humans are subject. Modern evolutionary theory by no means requires the answer yes. Many characteristics of organisms are mere byproducts of historical accident, and some characteristics are badly engineered for the purposes which they serve. An example is the mammalian eye, in which light has to pass through nerve fibers before it reaches light-sensitive tissue, and the optic nerve causes a blind spot at the point where it passes through the retina (Williams 1966; 1992). (Octopuses' eyes are more efficient from this point of view.) Natural selection can only tinker with what is genetically available, and perfect outcomes are often beyond its reach. So how much of grammar is well engineered, and how much of it is less than perfect owing to historical constraints?

Pinker and Bloom (1990), Newmeyer (1991), and Jackendoff (2002) are inclined to emphasize the positive aspects of grammatical engineering. That is understandable against the background of Chomsky's emphasis on neutral or even negative aspects. A different tack is taken by Carstairs-McCarthy (1999). He argues that the grammatical distinction between sentences and noun phrases, despite its familiarity and apparent inevitability, is in fact a piece of mediocre engineering, reflecting the cooption for syntactic purposes of neural mechanisms which evolved originally for the organization of the speech chain into syllables. He suggests that many of the syntactic habits of sentences, verbs, and noun phrases are reflections of the phonological habits of syllables, syllable nuclei (usually vowels), and syllable margins (consonants). This view is consistent with the proximity of Broca's area to the oral portion of the motor strip, as well as the frequent coincidence of grammatical and phonetic symptoms in Broca's aphasia. Carstairs-McCarthy also argues (2010) that some complexities of relationships between

word-forms, involving either affixes (as in *hands*, plural of *hand*) or internal changes (as in *feet*, plural of *foot*), are unexpected byproducts of cognitive factors that help young children to learn words quickly. But invoking imperfections in linguistic engineering as clues to the evolutionary origin of language is a new line of inquiry, and it remains to be seen how fruitful it will be.

7 Conclusion

This tour of recent work on the origins of language has revealed few solid, uncontroversial conclusions. Nevertheless, the field has entered an exciting period. The long freeze in relations between linguists and other language origin researchers has begun to thaw, just when discoveries in archeology, anthropology, primatology, and brain science are all helping to shed new light on the topic from a variety of directions. Will the evolution of language eventually come to be seen by linguistic theorists as not merely a quaint sideline but an essential source of evidence about why Universal Grammar is as it is? My guess is that it will, though the process may take a further decade or more. Certainly, the justification for the Paris Linguistic Society's ban no longer exists.

SUGGESTIONS FOR FURTHER READING

Aitchison, J. (1996). *The Seeds of Speech: Language Origin and Evolution*. Cambridge: Cambridge University Press.
A clear, easy-to-read introduction by a linguist, though its content is now somewhat dated.

Burling, R. (2005). *The Talking Ape: How Language Evolved*. Oxford: Oxford University Press.
Another easy and linguistically well-informed introduction.

Anderson, S. R. (2004). *Doctor Dolittle's Delusion: Animals and the Uniqueness of Human Language*. New Haven: Yale University Press.
A linguist's view of the similarities and differences between animal communication and human language.

Bickerton, D. (1990). *Language and Species*. Chicago: University of Chicago Press.
This item and the next are milestones in the recent revival of interest in language evolution. Neither is excessively technical.

Pinker, S. and Bloom, P. (1990). Natural language and natural selection [with peer commentary]. *Behavioral and Brain Sciences*, 13, 707–84.

Christiansen, M. H. and Kirby, S. (eds) (2003). *Language Evolution*. Oxford: Oxford University Press.
A survey of the state of the art around the beginning of the millennium, aimed at scholars rather than general readers, by a range of authors all of whom are respected researchers in the field.

Carstairs-McCarthy, A. (ed.) (2007). Special issue of *Lingua*: "The evolution of language." *Lingua* 117, issue 3.
A more recent state-of-the-art survey, focusing on contributions by linguists.

Jenkins, L. (ed.) (2004). *Variation and Universals in Biolinguistics*. Amsterdam: Elsevier.
An advanced collection of papers in the current Chomskyan "biolinguistic" tradition.

RELEVANT JOURNALS AND SOCIETIES

There is no journal specifically dedicated to language evolution, although this is one of the topics covered by *Interaction Studies* (published by Benjamins).

Since 1996 an International Conference on the Evolution of Language (EVOLANG) has been held every two years (http://www.ling.ed.ac.uk/evolang.org). Papers from these conferences have appeared in volumes published by Cambridge and Oxford University Presses, but there is no formal membership.

EMERGING TRENDS AND RESEARCH QUESTIONS

1. In the organization of the language faculty as it has evolved, where does the balance lie between aspects that are inevitable in some physical or mathematical sense (as emphasized by Chomsky), those that are products of adaptation through natural selection (as emphasized by Pinker and others), and those that are residues of historical accident (as emphasized by Carstairs-McCarthy)?
2. In its semantic organization, does language merely build on and elaborate cognitive capacities already present in apes, or have those cognitive capacities been radically altered by the syntax of human language?
3. How did the evolution of the capacity to vocalize interact with other elements in language evolution?
4. What was the role in language evolution of social factors such as grooming and sexual selection, as opposed to factors associated with knowledge of the environment (food sources, toolmaking and so on)?

REFERENCES

Aiello, L. C. (1996a). Hominine preadaptations for language and cognition. In P. Mellars and K. Gibson (eds), *Modelling the Early Human Mind* (pp. 89–99). Cambridge: McDonald Institute Monographs.

Aiello, L. C. (1996b). Terrestriality, bipedalism and the origin of language. In W. G. Runciman, J. M. Smith, and R. I. M. Dunbar (eds), *Evolution of Social Behaviour Patterns in Primates and Man* (pp. 269–89). (Proceedings of the British Academy 88.) Oxford: Oxford University Press.

Aiello, L. C. and Dean, C. (1990). *An Introduction to Human Evolutionary Anatomy*. London: Academic Press.

Arbib, M. A. (ed.) (2006). *Action to Language via the Mirror Neuron System*. Cambridge: Cambridge University Press.

Armstrong, D. F., Stokoe, W. C., and Wilcox, S. E. (1995). *Gesture and the Nature of Language*. Cambridge: Cambridge University Press.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight (eds), *Approaches to the Evolution of Language: Social and Cognitive Bases* (pp. 405–26). Cambridge: Cambridge University Press.

Berwick, R. C., Niyogi, P., and Jenkins, L. (1998). Language evolution and the Minimalist Program: the origins of syntax. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight (eds), *Approaches to the Evolution of Language: Social and Cognitive Bases* (pp. 320–40). Cambridge: Cambridge University Press.

Bickerton, D. (1981). *Roots of Language*. Ann Arbor, MI: Karoma.

Bickerton, D. (1990). *Language and Species*. Chicago: University of Chicago Press.

Bickerton, D. (1995). *Language and Human Behavior*. Seattle, WA: University of Washington Press.

Bickerton, D. (2002). Foraging versus social intelligence in the evolution of protolanguage. In A. Wray (ed.), *The Transition to Language* (pp. 207–25). Oxford: Oxford University Press.

Bradshaw, J. and Rogers, L. (1992). *The Evolution of Lateral Asymmetries, Language, Tool Use, and Intellect*. San Diego, CA: Academic Press.

Byrne, R. W. and Whiten, A. (eds) (1988). *Machiavellian Intelligence*. Oxford: Clarendon Press.

Calvin, W. H. (1993). The unitary hypothesis: a common neural circuitry for novel manipulations, language, plan-ahead, and throwing? In K. R. Gibson and T. Ingold (eds), *Tools, Language and Cognition in Human Evolution* (pp. 230–50). Cambridge: Cambridge University Press.

Calvin, W. H. and Ojemann, G. A. (1994). *Conversations with Neil's Brain: The Neural Nature of Thought and Language*. Reading, MA: Addison-Wesley.

Cann, R. L., Rickards, O., and Koji Lum, J. (1994). Mitochondrial DNA and human evolution: our one lucky mother. In M. H. Nitecki and D. V. Nitecki (eds), *Origins of Anatomically Modern Humans* (pp. 135–48). New York: Plenum.

Cann, R. L., Stoneking, M., and Wilson, A. L. (1987). Mitochondrial DNA and human evolution. *Nature*, 325, 31–6.

Carstairs-McCarthy, A. (1999). *The Origins of Complex Language: An Inquiry into the Evolutionary Beginnings of Sentences, Syllables and Truth*. Oxford: Oxford University Press.

Carstairs-McCarthy, A. (2010). *The Evolution of Morphology*. Oxford: Oxford University Press.

Cavalli-Sforza, L. L. and Cavalli-Sforza, F. (1995). *The Great Human Diasporas: The History of Diversity and Evolution*. Reading, MA: Addison-Wesley.

Cheney, D. L. and Seyfarth, R. M. (1990). *How Monkeys See the World: Inside the Mind of Another Species*. Chicago, IL: University of Chicago Press.

Chomsky, N. (1988). *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, MA: MIT Press.

Chomsky, N. (1991). Linguistics and cognitive science: problems and mysteries. In A. Kasher (ed.), *The Chomskyan Turn* (pp. 26–53). Oxford: Blackwell.

Christiansen, M. H. and Kirby, S. (eds) (2003). *Language Evolution*. Oxford: Oxford University Press.

Corballis, M. (2002). *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.

Deacon, T. W. (1997). *The Symbolic Species: The Coevolution of Language and the Brain*. New York: W. W. Norton.

Donald, M. (1991). *Origins of the Modern Mind*. Cambridge, MA: MIT Press.

DuBrul, E. L. (1958). *Evolution of the Speech Apparatus*. Springfield, IL: Charles C. Thomas.

Dunbar, R. (1996). *Grooming, Gossip and the Evolution of Language*. London: Faber.

Falk, D. (1992). *Braindance*. New York: Holt.

Fitch, W. T. (2002). Comparative vocal production and the evolution of speech: reinterpreting the descent of the larynx. In A. Wray (ed.), *The Transition to Language* (pp. 21–45). Oxford: Oxford University Press.

Gardner, R. A., Gardner, B. T., and Van Cantfort, T. E. (eds) (1989). *Teaching Sign Language to Chimpanzees*. Albany, NY: State University of New York Press.

Goodall, J. (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge, MA: Belknap Press.

Greenfield, P. M. (1991). Language, tools and the brain: the ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, 14, 531–51.

Hauser, M. D. (1996). *The Evolution of Communication*. Cambridge, MA: MIT Press.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298, 1567–79.

Hurford, J. R., Studdert-Kennedy, M., and Knight, C. (eds) (1998). *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge: Cambridge University Press.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jenkins, L. (ed.) (2004). *Variation and Universals in Biolinguistics*. Amsterdam: Elsevier.

Knight, W. (2005). Flores fossil passes unique species test. NewScientist.com news service, 12 March 2005.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., and Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, 413 (issue 6855), 519–23.

Lieberman, P. (1984). *The Biology and Evolution of Language*. Cambridge, MA: Harvard University Press.

Lieberman, P. and Crelin, E. S. (1971). On the speech of Neanderthal Man. *Linguistic Inquiry*, 11, 203–22.

Marcus, G. F., and Fisher, S. E. (2003). FOXP2 in focus: what can genes tell us about speech and language? *Trends in Cognitive Science*, 7, 257–62.

Newmeyer, F. J. (1991). Functional explanation in linguistics and the origins of language. *Language and Communication*, 11, 3–28.

Noble, W. and Davidson, I. (1996). *Human Evolution, Language and Mind: A Psychological and Archaeological Inquiry*. Cambridge: Cambridge University Press.

Pinker, S. (1994). *The Language Instinct*. New York: Morrow.

Pinker, S. and Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, 95, 201–36.

Pinker, S. and Bloom, P. (1990). Natural language and natural selection [with peer commentary]. *Behavioral and Brain Sciences*, 13, 707–84.

Poizner, H., Klima, E. S., and Bellugi, U. (1987). *What the Hands Reveal about the Brain*. Cambridge, MA: MIT Press.

Premack, D. (1976). *Intelligence in Ape and Man*. Hillsdale, NJ: Lawrence Erlbaum.

Ruhlen, M. (1994). *The Origin of Language: Tracing the Evolution of the Mother Tongue*. New York: John Wiley & Sons.

Savage-Rumbaugh, E. S. and Lewin, R. (1994). *Kanzi: The Ape at the Brink of the Human Mind*. New York: Wiley.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., and Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, serial no. 233, vol. 58, nos. 3–4.

Sommer, V. (1992). *Der Lob der Lüge*. Munich: Beck.

Steels, L. (1998). The synthetic modeling of language origins. *Evolution of Communication*, 1, 1–34.

Stringer, C. and Andrews, P. (2005). *The Complete World of Human Evolution*. London: Thames & Hudson.

Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua*, 117, 579–604.

Terrace, H. S. (1979). *Nim*. New York: Knopf.

Terrace, H. S. (2002). Serial expertise and the evolution of language. In A. Wray (ed.), *The Transition to*

Language (pp. 64–90). Oxford: Oxford University Press.

Trinkaus, E. and Shipman, P. (1993). *The Neandertals: Changing the Image of Mankind*. New York: Knopf.

Williams, G. C. (1966). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton, NJ: Princeton University Press.

Williams, G. C. (1992). *Natural Selection: Domains, Levels and Challenges*. New York: Oxford University Press.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, 18, 47–67.

Wray, A. (ed.) (2002). *The Transition to Language*. Oxford: Oxford University Press.

2 Languages of the World

BERNARD COMRIE

1 Introduction

The aim of this chapter is to provide readers with an overview of current views on the distribution of the languages of the world and on the genetic relations among those languages. Needless to say, the mention of individual languages will be on a selective basis, with emphasis on those languages that are most widely spoken or that have played an important role in history, although some departure from this principle will necessarily be made for parts of the world, like the Americas, Australia, and New Guinea, where there are few languages with large numbers of speakers.

The best currently available detailed account of the distribution of the world's languages, with information on geographic location, number of speakers, and genetic affiliation, is Lewis, Simon, and Fennig (2016), also available online at https://ethnologue.com. This work lists over 6,700 languages spoken in the world today or having recently become extinct. While this figure is toward the high end of estimates that would be given by linguists, it is nonetheless a reasonable estimate, based where possible on a linguists' definition of "language" (as opposed to "dialect") as a speech variety that is not mutually intelligible with other speech varieties. This definition brings with it a number of ancillary problems. For instance, testing mutual intelligibility is far from straightforward (Casad 1974). There are, moreover, complicated cases, like intelligibility that is greater in one direction than the other, i.e. speakers of A understand B better than speakers of B understand A, and dialect chains, i.e. a geographic chain of dialects A—B— … —N such that each dialect is mutually intelligible with its neighbor(s), but the extremes of the chain, A and N, are not mutually intelligible. Added to this is the fact that for many speech varieties serious tests of mutual intelligibility have simply not been carried out.

The question of the genetic affiliation among the languages of the world is one that is currently fraught with controversy, in particular between those who adopt a cautious stance, accepting that languages are genetically related only in the face of overwhelming evidence, and those who are more willing to accept genetic relatedness among languages on the basis of less compelling evidence. In this survey I have in general included only language families that are universally or almost universally recognized by linguists, and I have specifically added notes of caution where I use terms that cover larger potential genetic groupings of languages. At the same time, this survey does have a duty to inform the reader about more speculative hypotheses that have gained the support of a reasonable set of linguists. I have therefore included notes on possible more widespread groupings, largely following Ruhlen (1994: 15–34). The other book dealing with the classification of the world's languages, Ruhlen (1987), likewise adopts in general an approach that includes both traditionally accepted and currently debated genetic groupings, though with

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

an equally undisguised bias toward the latter; but the critical approach to less widely accepted groupings does not extend to its treatment of languages of the Americas.

In the space available I have limited myself to geographic distribution and genetic affiliation, although there are a number of other questions that might have been touched on, such as the ways in which languages influence one another by contact, and more generally the historical processes that have given rise to the present-day distribution of the world's languages. Recent literature dealing with this latter problem includes Nichols (1992), Dixon (1997), and Nettle (1999).

Many of the references for individual language families are to volumes in the following series: Cambridge Language Surveys (Cambridge University Press, ongoing), Routledge Language Family Descriptions (Routledge, discontinued and effectively replaced by the following), Curzon Language Family Descriptions (Curzon, ongoing).

2 Languages of Europe and Northern Asia

2.1 Indo-European languages

The Indo-European language family (Ramat and Ramat 1998) covers most of Europe and spreads, with some breaks, across Iran and Central Asia down into South Asia. As a result of colonial expansion, it is now also dominant in the Americas and in Australia and New Zealand. In Europe itself, only a few peripheral areas are occupied by non-Indo-European languages, in particular areas where Basque and some Uralic languages are spoken and parts of the Caucasus. The Indo-European family subdivides into a number of well-established branches.

The Germanic languages (König and van der Auwera 1994) are the dominant languages of northwestern Europe, extending into central Europe. This is the language family that includes English, and also Dutch, German, and the Scandinavian languages (including Danish, Norwegian, Swedish, and Icelandic); an offshoot of German with considerable admixture from Hebrew-Aramaic and Slavic is Yiddish, the traditional language of Ashkenazi Jews and a widely spoken language of eastern Europe before the Holocaust. The Scandinavian languages form North Germanic, while the other languages cited are West Germanic; a third subbranch of the family, East Germanic, is now extinct, the only substantially attested language being Gothic.

The Celtic languages (Ball 1993; MacAulay 1993) were once also dominant languages of western and central Europe but, with the expansion of Germanic and Romance languages in particular they have retreated to the western fringes of Europe, the living languages being Welsh in Wales, Irish on the west coast of Ireland, Breton in Brittany (France), and Scots Gaelic in northwestern Scotland.

The Romance languages (Harris and Vincent 1988; Posner 1996) occupy most of southwestern Europe, and are the descendants of Latin, the language of the Roman Empire. Strictly speaking, the branch of Indo-European is Italic, since it includes a number of languages other than Latin that died out by the early centuries of the Common Era as a result of Roman and Latin expansion, so that all living Italic languages are in fact Romance languages. The major living languages are French, Catalan, Spanish, Portuguese, Italian, and Romanian.

Turning to eastern Europe, the northernmost Indo-European branch is Baltic, now consisting of the two languages Lithuanian and Latvian. The Baltic languages have a particularly close relation to the Slavic (Slavonic) languages (Comrie and Corbett 1993), now dominant in much of eastern and central Europe and including three subbranches. The East Slavic languages are Russian, Belarusian (Belorussian), and Ukrainian. The West Slavic languages include Polish, Czech, and Slovak. The South Slavic languages are Slovenian, Serbo-Croatian, Bulgarian, and Macedonian. As a result of ethnic differences, what linguists would, on grounds of mutual intelligibility, consider a single Serbo-Croatian language is now often divided into Serbian and Croatian, with Bosnian sometimes added as a third ethnic variety.

Two further branches of Indo-European, each consisting of a single language, are found in the Balkans. Albanian consists of two dialect groups, Gheg in the north and Tosk in the south, which

might well be considered distinct languages on the basis of the mutual intelligibility test, although there is a standard language based on Tosk. Hellenic includes only Greek, although it is customary to give a different name to the branch, in part because it includes varieties of Greek over more than three millennia, from Mycenean through Classical Greek and Byzantine Greek to the modern language. Armenian, spoken primarily in Armenia though also in the Armenian diaspora originating in eastern Turkey, is another branch of Indo-European consisting of a single language, although the differences between Eastern Armenian (spoken mainly in Armenia) and Western Armenian (spoken originally mainly in Turkey) are considerable, and there are two written languages.

Finally, with respect to the living languages, the Indo-Iranian languages are spoken from the Caucasus to Bangladesh. Indo-Iranian divides into two subbranches, Iranian and Indo-Aryan (Indic), the latter occupying an almost continuous area covering most of Pakistan, northern India, Nepal, and Bangladesh. The most widely spoken Iranian languages are Persian (Iran), with national variants Tajik (in Tajikistan) and Dari (in Afghanistan), Kurdish (mainly in the border area of Turkey, Iran, and Iraq), Pashto (in Afghanistan and Pakistan), and Balochi (in Pakistan).

The Indo-Aryan subbranch of Indo-Iranian (Masica 1991) includes Sanskrit, the classical language of Indian civilization; Pali, the sacred language of Buddhism; and a large number of modern languages, of which the most widely spoken are Hindi and Urdu, essentially different national forms of the same language, in India and Pakistan respectively; Sindhi and Western Panjabi (Lahnda) in Pakistan; Nepali in Nepal; and Kashmiri, Eastern Panjabi, Gujarati, Rajasthani, Marathi, Bhojpuri, Maithili, Assamese, and Oriya in India; Bengali in India and Bangladesh; and Sinhala, geographically separated from the other Indo-Aryan languages in Sri Lanka. It should also be noted that the various Romani languages, spoken by Rom (Gypsies), belong to the Indo-Aryan group of languages.

In addition, two branches of Indo-European consist of extinct but well-attested languages. The best known of the Anatolian languages, spoken in what is now Turkey, is Hittite, language of a major ancient empire (seventeenth to twelfth centuries BCE). Tocharian is a family of two closely related languages, attested in texts from the latter half of the first millennium CE in what is now the Xinjiang region in northwestern China.

2.2 *Uralic languages*

The Uralic language family (Abondolo 1998) must once have been spoken over a continuous part of northeastern Europe and northwestern Asia, but inroads by other languages, primarily Indo-European and Turkic, have isolated many of the Uralic branches and languages from one another geographically. The family falls into two clear subgroups, Finno-Ugric and Samoyedic. The Samoyedic languages, all with small numbers of speakers, are spoken along the northern fringe of Eurasia, roughly from the Kanin peninsula to the Taymyr peninsula.

Finno-Ugric divides in turn into a number of branches: Balto-Finnic (around the Baltic Sea), Saamic (Lappish) (northern Scandinavia to the Kola peninsula), Volgaic (on the Volga, although the unity of this branch is now questioned), Permic (northeastern European Russia), and Ugric (western Siberia and Hungary, though the unity of Ugric is also questioned). The most widely spoken languages are two Balto-Finnic languages, Finnish and Estonian, and one of the Ugric languages, Hungarian. It should be noted that the present location of Hungarian is the result of a long series of migrations, so that Hungarian is now far distant in location from its closest relatives within Finno-Ugric.

2.3 *Altaic families*

Altaic is a proposed genetic grouping that would include minimally the Turkic, Tungusic, and Mongolic families, perhaps also Korean and Japanese. Each of these components is a

well-established language family, and Altaic lies perhaps at the dividing line that separates proponents of wide-ranging genetic groupings of languages from those that remain skeptical. Here the various families and the languages they contain will be noted without any commitment to the unity of the overall grouping.

The Turkic languages (Johanson and Csató 1998) are spoken, with interruptions, in a broad belt stretching from the Balkans in the west through the Caucasus and Central Asia and into Siberia. Classification of the Turkic languages has always been problematic, in part because most of the languages are very close to one another linguistically, in part because population movements and even, in recent times, language politics have tended to overlay new distinctions on old ones. It is recognized that two languages form separate branches of the family: Chuvash, spoken in the Chuvash Republic (Russia) on the Volga, and Khalaj, spoken by a small and dwindling population in the Central Province of Iran. Johanson and Csató (1998: 82–3) propose four other branches, listed here with representative languages. Southwestern (Oghuz) Turkic includes Turkish (Turkey), Azeri (Azerbaijani) (Azerbaijan, northwestern Iran), and Turkmen (Turkmenistan, also Iran and Afghanistan). Northwestern (Kipchak) Turkic includes Kumyk and Karachay-Balkar (both spoken in the Caucasus), Tatar and Bashkir (both spoken on the Volga), Kazakh (Kazakhstan and northwestern China), and Kirghiz (Kyrgyzstan). Southeastern (Uyghur) Turkic includes Uzbek (Uzbekistan) and Uyghur (mainly in northwestern China). Finally, Northeastern (Siberian) Turkic includes Tuvan and Altai (Oyrot) in southern Siberia and Yakut (Sakha) in the huge Sakha Republic in Russia.

The Tungusic languages have few speakers, scattered across the sparsely populated areas of central and eastern Siberia, including Sakhalin Island, and adjacent parts of northeastern China and Mongolia. One Tungusic language, Manchu, is well known in history as the language of the Manchu conquerors who established the Qing dynasty in China (1644–1911), but all but a few ethnic Manchu now speak Mandarin.

The Mongolic languages are spoken primarily in Mongolia and adjacent parts of Russia and China, although there is also one Mongolic language in Afghanistan while Kalmyk is spoken in Kalmykia (Russia) on the lower Volga. The most widely spoken Mongolic language is Mongolian (Mongolia, northern China), although both Buriat (to the south and east of Lake Baikal) and Kalmyk are languages of constituent republics of the Russian Federation.

The other two potential members of the Altaic family are Korean and Japanese. Korean (Sohn 1999) is a single language. Japanese (Shibatani 1990) is strictly speaking a small family, including not only Japanese but also the Ryukyuan languages, which are not mutually intelligible with Japanese or with each other; the family is sometimes called Japanese-Ryukyuan.

2.4 Chukotko-Kamchatkan languages

Chukotko-Kamchatkan is a small language family spoken on the Chukotka and Kamchatka peninsulas in the far northeast of Russia. All of the languages, which include Chukchi, are endangered.

2.5 Caucasian families

Some of the languages spoken in the Caucasus belong to language families already mentioned, in particular Indo-European (Armenian, Iranian) and Turkic. But there remain a large number of languages that do not belong to any of these families. These languages are referred to as Caucasian, but it is important to note that this is essentially a negative characterization. Indeed, it is currently believed that there are two or three families represented among the "Caucasian" languages.

The Kartvelian (South Caucasian) family is spoken in Georgia with some extension into Turkey, and the main language, the only one to be used as a written language, is Georgian, the official language of the Republic of Georgia.

The other two Caucasian families are Northwest Caucasian (West Caucasian, Abkhaz-Adyghe) and Northeast Caucasian (East Caucasian, Nakh-Daghestanian), although Nikolayev and Starostin (1994) present a detailed argument for considering them to constitute a single North Caucasian family; I will treat them separately here.

The Northwest Caucasian languages are spoken in Abkhazia, the northwestern part of the geographic territory of the Republic of Georgia, and in parts of Russia to the north of this. The main languages are Abkhaz (in Abkhazia) and the varieties of Circassian (Kabardian and Adyghe) spoken in Russia and by a sizeable diaspora in the Middle East.

The Northeast Caucasian languages are spoken primarily in the constituent republics of the Russian Federation of Chechnya, Ingushetia, and Daghestan, with some spillover into Azerbaijan. The languages with the largest numbers of speakers are Chechen (Chechnya) and Avar (Dagestan).

2.6 Other languages of Europe and northern Eurasia

A number of other languages or small language families are or were spoken in Europe or northern Asia but do not, at least unequivocally, belong to any of the above families. Basque is a language isolate spoken in the Pyrenees, divided by the Spain–France border. Etruscan was the language of Etruria in northern Italy before the spread of Latin; it is now known to be related to two less well-attested languages, Rhaetian in the Alps and Lemnian on the island of Lemnos (Limnos) in the Aegean. Hurrian (sixteenth century BCE) and Urartean (ninth to seventh centuries BCE) are two related extinct languages once spoken in eastern Anatolia.

The Yeniseian family of languages has only one survivor, Ket, spoken on the Yenisei River in western Siberia, although other languages are known from historical records that became extinct from the eighteenth to the twentieth centuries. Yukaghir, spoken in the area of the Kolyma and Indigirka rivers in northeastern Russia, is sometimes treated as a language isolate, although many linguists believe that it is distantly related to Uralic. Nivkh (Gilyak) is a language isolate spoken at the mouth of the Amur River and on Sakhalin Island. Ainu (Shibatani 1990) is a virtually extinct language isolate spoken in northern Japan (Hokkaido Island). Some or all of the languages mentioned in this paragraph are often referred to collectively as Paleosiberian or Paleoasiatic, but this is essentially a negative characterization (they do not belong to any of the established language families), with no implication that they are related to one another.

2.7 Proposals for larger groupings

Two similar, but not identical, proposals have been made for grouping together a large number of the language families found in Europe and northern Asia. The Nostratic proposal, first worked out in detail by Illič-Svityč (1971–84), would include at least Indo-European, Uralic, Altaic, Afroasiatic (see Section 4.1), Kartvelian, and Dravidian (see section 3.1). Eurasiatic, the subject of work by Joseph H. Greenberg, would include at least Indo-European, Uralic, Altaic, Chukotko-Kamchatkan, Eskimo-Aleut (see Section 5.1), and possibly also Nivkh. For possibilities including some of the other languages, see Section 3.10.

3 Languages of Southern, Eastern, and Southeastern Asia and Oceania

This section deals primarily with languages of southeast Asia and its island extensions into Oceania. James Matisoff (2002) has produced a general survey of southeast Asia. Things are

somewhat better for the islands, although this is an area where there is rapid ongoing work leading to frequent changes in accepted genetic classification.

3.1 Dravidian languages

The Dravidian languages (Steever 1998) are the dominant languages of southern India, with Tamil also spoken in northern Sri Lanka. The Dravidian family is divided into four branches, Northern, Central, South-Central, and Southern, although the four main, literary languages belong to the last two branches. Telugu, the language of the Indian state of Andhra Pradesh, is a South-Central Dravidian language, while the following are South Dravidian: Tamil (Tamil Nadu state in India, northern Sri Lanka), Malayalam (Kerala state in India), and Kannada (Karnataka state in India).

3.2 Austro-Asiatic languages

Austro-Asiatic languages are spoken from eastern India across to Vietnam and down to the Nicobar Islands and peninsular Malaysia, although in most of this region they are interspersed among other, more widely spoken, languages. The family has two branches, Munda and Mon-Khmer. Munda languages are spoken in eastern India and some neighboring regions. Most of the languages have small numbers of speakers, the main exceptions being Santali and Mundari. Mon-Khmer languages start in eastern India, but their largest numbers are in Myanmar, Thailand, Malaysia, Cambodia, Laos, and Vietnam. While most have few speakers, there are two notable exceptions. Khmer (Cambodian) is the dominant language of Cambodia, while Vietnamese is the dominant language of Vietnam. Another historically important Mon-Khmer language is Mon, still spoken in the delta area to the east of Yankon (Rangoon), as the Mon played an important role in the development of Burmese and Thai culture. Vietnamese is typologically quite unlike the other Mon-Khmer languages and has undergone considerable influence from Chinese, with the result that its membership in Mon-Khmer was for a long time not recognized.

3.3 Sino-Tibetan

Sino-Tibetan is one of the world's largest language families in terms of numbers of speakers, and includes the language most widely spoken as a native language, namely Mandarin Chinese. Sino-Tibetan languages are spoken primarily in China, the Himalayan region of India and Nepal, and Myanmar, with excursions into some neighboring countries, in addition to a large Chinese diaspora. (Ethnic Chinese make up, for instance, some three-quarters of the population of Singapore.) Our understanding of Sino-Tibetan has been increased considerably in recent years by the availability of descriptions of the less widely spoken languages; a major impetus here has been the Sino-Tibetan Etymological Dictionary and Thesaurus project (see http://www.linguistics.berkeley.edu/lingdept/research/stedt/) at the University of California at Berkeley.

The usual classification splits the family into two branches, Sinitic (consisting essentially of the Chinese languages; Norman 1988) and Tibeto-Burman. Recently, van Driem (1997) has proposed, on the basis of the most recent reconstructions of the phonology of Old Chinese, that Sinitic may actually be a subbranch of Tibeto-Burman, grouped most closely with the Bodic languages – the family as a whole would thus more properly be called Tibeto-Burman. In what follows I will retain the traditional classification, though emphasizing that this is more for convenience than through conviction.

Sinitic consists primarily of the various Chinese languages, which in terms of mutual intelligibility are clearly sufficiently different from one another to be considered distinct languages, even if all stand under the umbrella of a reasonably homogeneous written language. The major

varieties are Mandarin, Wu (including Shanghai), Gan, Hakka, Xiang, Yue (Cantonese), Northern Min, and Southern Min (including Taiwanese).

The main groupings within (traditional) Tibeto-Burman are Baric, Bodic, Burmese-Lolo, Karen, Nung (Rung), and Qiang; proposals for subgrouping vary. The Baric languages include Meithei (Manipuri) in Manipur State, India. Bodic includes a number of languages spoken in the Himalayas, the most widely spoken and culturally important being Tibetan. The Burmese-Lolo languages are spoken mainly in Myanmar and southern China and include Burmese. The Karen languages are spoken in Myanmar and adjacent parts of Thailand, the most widely spoken being S'gaw Karen (White Karen). The Nung and Qiang languages are spoken in Myanmar and southern China.

3.4 Daic languages

Daic is one of a number of names (others including Tai-Kadai and Kam-Tai) for a family of languages with three branches, Kadai, Kam-Sui, and Tai. Kadai and Kam-Sui contain languages with small numbers of speakers spoken in southern China and parts of Vietnam. Tai, by contrast, includes two of the dominant languages of southeast Asia, namely Thai (Thailand) and the closely related Lao (Laos). Other Tai languages are spoken in these countries and in southern China, though with some excursions into Vietnam and Myanmar. The most widely spoken Tai language of China is Zhuang. It is now conventional to use the spelling Thai for the language, Tai for the branch, and Daic for the family.

3.5 Hmong-Mien (Miao-Yao) languages

The Hmong-Mien or Miao-Yao languages are spoken in parts of southern China and stretching into southeast Asia, especially Vietnam. Hmong and Mien are the indigenous ethnic names, while Miao and Yao are the Chinese equivalents. Hmong and Mien are the two branches of the family, and each consists of several languages. The most widely spoken variety is Hmong Njua (Western Hmong) in China and Vietnam.

3.6 Austronesian languages

Austronesian is one of the most extensive families, covering almost all the islands bounded by an area from Madagascar in the west via Taiwan and Hawaii to Easter Island in the east and down to New Zealand in the south, with the exception of most of New Guinea and all of Australia. Although predominantly an island language family, Austronesian languages are also dominant in peninsular Malaysia, while the Chamic languages are spoken in coastal areas of Vietnam and Cambodia as well as on Hainan Island, China. An overview of the Austronesian languages by Robert Blust (2013) is available.

Although the Austronesian languages of Taiwan are very much minority languages on an island where varieties of Chinese have become dominant, the internal diversity among the Austronesian languages of Taiwan, the so-called Formosan languages, is greater than that in all the rest of Austronesian put together, so there is a major genetic split within Austronesian between Formosan and the rest, the latter now usually called Malayo-Polynesian (although in some earlier work this term was used for the family as a whole). Indeed, the genetic diversity within Formosan is so great that it may well consist of several primary branches of the overall Austronesian family.

The basic internal classification of Malayo-Polynesian is reasonably well established. The primary branchings are into Western Malayo-Polynesian and Central-Eastern Malayo-Polynesian, with the dividing line running to the east of Sulawesi and through the Lesser Sunda islands. Western Malayo-Polynesian thus includes all the languages of the Philippines, the Asian

mainland, western Indonesia, and Madagascar. It also includes all the Austronesian languages with large numbers of speakers, including Malay–Indonesian, the different national varieties of what is essentially the same standard written language, though with radically different local spoken varieties. Other widely spoken languages of Indonesia are Acehnese, Toba Batak, Lampung, and Minangkabau (all on Sumatra), Javanese, Madurese, and Sundanese (all on Java), Balinese (on Bali), and Buginese and Makassarese (on Sulawesi). Widely spoken languages of the Philippines include, in addition to the national language Tagalog, the following: Bikol, Hiligaynon, Ilocano, Pampangan, Pangasinan, and Waray-Waray. The other major Western Malayo-Polynesian language is Malagasy (Madagascar).

Central-Eastern Malayo-Polynesian further divides into Central Malayo-Polynesian and Eastern Malayo-Polynesian, the former comprising a number of languages spoken in parts of the Lesser Sunda islands and of southern and central Maluku. Eastern Malayo-Polynesian divides in turn into South Halmahera-West New Guinea and Oceanic, with the former including Austronesian languages of southern Halmahera and parts of northwest Irian Jaya. Oceanic includes all other Austronesian languages of Melanesia, Micronesia (except that Palauan and Chamorro are Western Malayo-Polynesian), and Polynesia. Oceanic thus includes the Polynesian languages, spoken in the triangle whose points are Hawaii in the north, Easter Island in the east, and New Zealand in the south. Polynesian languages include Hawaiian, Tahitian, Maori, Samoan, Tuvaluan, and Tongan. Genetically just outside Polynesian within Oceanic is Fijian. Kiribati (Gilbertese) is a Micronesian language, also within Oceanic but outside Polynesian.

3.7 Papuan families

The island of New Guinea and immediately surrounding areas form the linguistically most diverse area on earth, with over 1,000 languages spoken by a population of between six and seven million. While some of the coasts of New Guinea itself and most of the smaller islands of the New Guinea area are occupied by Austronesian languages, most of the interior, together with some coastal and island areas, are occupied by so-called Papuan languages. The term "Papuan" is basically defined negatively as those languages of the New Guinea area that are not Austronesian. Until recently, two radically different approaches to the internal classification of Papuan languages prevailed among specialists. On the one hand, Wurm (1982) divided the languages into five major "phyla" (i.e. large-scale families) and six minor phyla, plus seven or more language isolates. The most widespread of these large families is the Trans New Guinea phylum, containing most of the languages spoken across the highland backbone of the island but also extending southwest as far as Timor and neighboring islands. The other major phyla in this classification are: West Papuan (northern Halmahera and parts of the Bird's Head in Irian Jaya), Geelvink Bay (part of the north coast of Irian Jaya, to the east of the Bird's Head), Torricelli (western parts of the north coast of Papua New Guinea), Sepik-Ramu (large parts of northwestern Papua New Guinea), and East Papuan (on islands from New Britain eastward to the Solomons). (Note that Geelvink Bay is now called Cenderawasih Bay; the Bird's Head was formerly called the Vogelkop.) Foley (1986), by contrast, maintains that work to date allows only the identification of about sixty genetic units, with internal diversification about as for Romance, among the Papuan languages, with higher level relations among them remaining a task for future research.

Ongoing work, some of it published in Pawley (1998) and including contributions by Foley among others, suggests that there may well be a firm basis for using traditional comparative methods for a stripped down version of the Trans New Guinea family, which would still include a substantial number of the smaller genetic units found along the backbone of the main island, although by no means all of Wurm's Trans New Guinea phylum finds justification in the ongoing work. But this does indicate that the time may be ripe or nearly ripe for a more systematic look at genetic relations among the Papuan languages.

As can be imagined from the low average ratio of speakers to languages, most Papuan languages have few speakers. The languages listed by Grimes (1996a) as having more than 100,000 speakers are Enga, Chimbu, and Medlpa in the highlands of Papua New Guinea, and Western Dani, Grand Valley Dani, and Ekari in the highlands of Irian Jaya. It is a general pattern that languages with more speakers tend to be found in the highlands, whose valleys are also the area of greatest population density.

3.8 Australian families

The classification of Australian languages is in something of a turmoil at present. Dixon (1980) proposed that all Australian languages form a single family, with the exception of Tiwi, spoken on islands off the north coast, and Djingili, in the Barkly Tableland. In a more recent work, Dixon (1997) takes a different stand, suggesting that the peculiar social history of Aboriginal Australia, with the absence of major power centers and continual contact among languages, may make the traditional comparative method unworkable for Australia. Many Australianists nonetheless retain the concept of language family, with about twenty language families in Australia, perhaps all or most being related as a single Australian language family. In particular, there is widespread acceptance of a Pama-Nyungan family that would include the languages spoken in most of the island-continent except some of those in the far north, although Dixon (1997) explicitly rejects the genetic unity of Pama-Nyungan. A new synthesis of Australian languages, to replace Dixon (1980), is currently in preparation by Dixon and others. No Australian language has a large number of speakers, the most viable languages having at most a few thousand.

The records of the extinct Tasmanian languages are sparse, and Dixon (1980) concludes that they are insufficient to exclude the possibility that they may have been related to Australian languages, though equally they are insufficient to establish such a relationship (or any other). Speakers of the Tasmanian languages must have been separated from the rest of humanity for about 12,000 years, from the time rising waters created the Bass Strait to the first visits by Europeans, making them the most isolated human group known to history; the genocide visited upon the Tasmanians in the nineteenth century is thus also a scientific tragedy of the first order.

3.9 Other languages of southern, eastern, and southeastern Asia

A number of living languages spoken in this region have so far eluded genetic classification, in particular Burushaski spoken in northern Pakistan, and Nahali (Nihali) in central India. Burushaski is reasonably well described, while Nahali is in urgent need of a detailed description. The Andamanese languages, spoken on the Andaman Islands (politically part of India) also lack any widely accepted broader genetic affiliation.

In addition, reference may be made to two extinct languages. Elamite was the language of Elam, an important empire in what is now southwestern Iran around 1000 BCE; it is possible that it may be related to Dravidian (McAlpin 1981). Sumerian was the language of ancient Sumer, and is noteworthy as being probably the first language to have had a writing system; it was still used as a literary language in the Old Babylonian period, although before or during this period it was replaced as a spoken language by Akkadian (see Section 4.1).

3.10 Proposals for larger groupings

For the suggestion that Dravidian might belong to the proposed Nostratic macrofamily, see Section 2.7.

Benedict (1975), building largely on his own earlier work, proposes an Austro-Tai macrofamily that would include Austro-Asiatic, Daic, Hmong-Mien, and Austronesian. Ruhlen

(1994: 24–8) reports on attempts to set up a Dene-Caucasian grouping that would include Na-Dene (see Section 5.1), Yeniseian, Sino-Tibetan, Nahali, Sumerian, Burushaski, North Caucasian, and Basque (for some of these languages, see Section 2.6).

Greenberg (1971) proposed an Indo-Pacific grouping that would include all Papuan languages plus the Andamanese and Tasmanian languages, but this proposal does not seem to have been taken up in detail by other linguists.

The possibility of a link between (some) Australian and (some) Papuan languages is mooted by Foley (1986).

4 Languages of Africa and Southwestern Asia

The starting point for recent discussions of the classification of African languages is Greenberg (1963), who proposes a fourway division into Afroasiatic, Niger-Congo (Niger-Kordofanian), Nilo-Saharan, and Khoisan families. Afroasiatic and Niger-Congo are now generally accepted, while more controversy has surrounded Nilo-Saharan and Khoisan.

4.1 Afroasiatic languages

The Afroasiatic (formerly Hamito-Semitic) family is the dominant language family of most of north Africa and large parts of southwestern Asia and, although individual languages have contracted or extended their geographical distribution, this distribution of the family as a whole goes back to antiquity. The family is generally considered to have six branches: Semitic in southwestern Asia, Eritrea, and much of Ethiopia, also of course now in most of North Africa as a result of the spread of Arabic; Egyptian in older times in Egypt; Berber across most of the rest of north Africa (though now in retreat before Arabic in most of this area); Chadic, in a belt centered on northern Nigeria and southern Niger; Cushitic in the Horn of Africa (Somalia, Djibouti, much of southern Ethiopia, and extending into Kenya and Tanzania to the east of Lake Victoria); and Omotic along the Omo River in southeastern Ethiopia. Omotic languages were formerly, and are still sometimes, considered a subbranch of Cushitic. There is need for an up-to-date survey of the family as a whole; in the meantime, reference may be made to Diakonoff (1988).

The Semitic languages are the best studied of the Afroasiatic branches, and Semitic languages can be traced back almost to the beginning of written history. The most recent survey is Hetzron (1997). The Semitic branch is divided into two subbranches, East Semitic and West Semitic. The East Semitic branch is extinct, although it contains Akkadian, the language of the Babylonian and Assyrian civilizations. West Semitic contains all the living Semitic languages as well as several historically important dead languages. The subdivision of West Semitic is more controversial, especially as regards the position of Arabic. The widely accepted current classification as given in Hetzron (1997) divides West Semitic into Central Semitic and South Semitic. Central Semitic subdivides into Arabic and Northwest Semitic. The older classification would put Arabic in South Semitic, and thus use Northwest Semitic for the other subbranch of West Semitic. The classification of Hetzron will be followed in the presentation here.

Arabic was, until the spread of Islam, the language of part of the Arabian peninsula, but as the language of Islam it has spread through much of southwestern Asia and north Africa, replacing the languages previously spoken across most of this area and becoming one of the modern world's major languages. The standard written language is still firmly rooted in the language of the Qur'ān and medieval Arabic literature, but spoken varieties of Arabic are sufficiently different from one another that mutual intelligibility is not possible between extreme varieties. However, only one variety of Arabic has developed as a separate written language, namely Maltese.

Northwest Semitic includes the Canaanite languages and Aramaic. The best known of the Canaanite languages is Hebrew, used as the spoken and written language of the Jews until the early centuries CE, then as a written and liturgical language by Jews throughout the middle ages, to be revived as a spoken language starting in the late nineteenth century and reaching its culmination as an official language and the dominant spoken language of Israel. The other Canaanite languages are all extinct, the best known being Phoenician. Aramaic was a major lingua franca of the Near East from the eighth century BCE, but at present varieties of Aramaic are spoken in enclaves in Syria, Iraq, and Iran.

South Semitic includes the South Arabian languages spoken on the southern fringe of the Arabian peninsula. Most living South Semitic languages belong to the Ethiopian Semitic subgroup, and include Amharic, the dominant language of Ethiopia; Tigrinya, an important regional language of Ethiopia and Eritrea; and Tigré, another regional language of Eritrea. In addition, Ethiopian Semitic includes Ge'ez, the extinct language still used liturgically by the Ethiopian Church.

Egyptian, by which is meant here Ancient Egyptian, is a single language attested in various historical stages from the earliest writing in Egypt. The hieroglyphic writing system and its offshoots were used into the Common Era, but were soon replaced after Christianization by a Greek-based script, and this later variety of the language is called Coptic. Coptic survived as a spoken language to the late middle ages, when it was finally replaced completely by Arabic, although it continues in use as the liturgical language of the Coptic Church. A recent survey is Loprieno (1995).

The Berber languages are spoken in a scattered pattern across north Africa from just east of the Egypt–Libya border, though they are strongest in mountainous parts of Algeria and especially Morocco, and in the desert parts of Mali and Niger. Among the most widely spoken varieties are Kabyle (Algeria), Chaouia (Algeria), Tarifit (Northern Shilha) (Morocco, Algeria), Tachelhit (Central Shilha) (Morocco, Algeria), Tamazight (Southern Shilha) (Morocco, Algeria), Tamashek (the language of the Tuaregs, mainly in Mali and Niger).

Most of the Chadic languages have few speakers, but there is one significant exception, namely Hausa, the dominant indigenous language of northern Nigeria and southern Niger. Hausa is widely used as a lingua franca by speakers of other neighboring Chadic and non-Chadic languages.

The most widely spoken Cushitic languages are Somali (mainly in Somalia and Ethiopia), Sidamo (Ethiopia), Oromo (Galla) (Ethiopia), Afar (Ethiopia, Eritrea, Djibouti), and Bedawi (Beja) (Sudan). The most widely spoken Omotic language is Wolaytta (Ethiopia).

4.2 Niger-Congo languages

Niger-Congo languages cover most of Africa south of a line drawn from the mouth of the Senegal River in the west to where the equator cuts the coast of Africa in the east, with the major exception of the area in southwestern Africa occupied by the Khoisan languages. There are also considerable excursions of Niger-Congo to the north of this line, and less significant excursions of non-Niger-Congo languages to the south of this line, e.g. Cushitic and Nilotic languages spoken to the east of Lake Victoria. The internal structure of the Niger-Congo family was first worked out in detail in Greenberg (1963), although a number of changes have been proposed in more recent work, several of which are still the subject of debate. The most recent overview is Bendor-Samuel (1989), and the classification given there will be followed here.

One branch of Niger-Congo is spoken outside the area delimited above, namely Kordofanian, spoken in the Nuba mountains of Sudan, to the south of El-Obeid. While Greenberg considered Kordofanian genetically the most distant of the languages in the overall family, thus naming the family as a whole Niger-Kordofanian with two coordinate branches Kordofanian and Niger-Congo, the current view is rather that Kordofanian is at least no more distant genetically from the core of the family than are the Mande languages, and the name Niger-Congo is current

for the family as a whole. It should be noted that one group of languages assigned tentatively by Greenberg to Kordofanian on the basis of fragmentary material, namely Kado or Kadu (formerly called Kadugli-Krongo), is now believed not to be Kordofanian or Niger-Congo, and perhaps Nilo-Saharan (Bender 1997: 25).

The Mande languages are spoken over most of west Africa to the west of 5°W and to the south of 15°N, although considerable parts of this territory, especially near the coasts, are occupied by other branches of Niger-Congo (Atlantic and Kru). Mande languages include Bambara, the major indigenous language of Mali, and some closely related languages such as Maninka; Jula, spoken in Côte d'Ivoire and Burkina Faso; Kpelle, the major indigenous language of Liberia; and Mende, the major indigenous language of Sierra Leone.

In Bendor-Samuel (1989: 21) the rest of Niger-Congo, once Kordofanian and Mande have been removed, is referred to as Atlantic-Congo, with Atlantic and Ijoid as the genetically next most divergent groups, the remainder being referred to as Volta-Congo. The remaining Niger-Congo groups, i.e., Bendor-Samuel's Atlantic-Congo, will be treated together in what follows.

Atlantic languages are spoken, predominantly in coastal areas, from the Senegal River in the north down into Liberia, although the most widely spoken Atlantic language, Fula (Fulfulde, Peul) has a different distribution. The Fulani, as the speakers of Fula are called, are pastoralists whose range is between the rain forest to the south and the desert to the north, with traditional seasonal moves along a north–south axis; the language is spoken in pockets from the Atlantic coast into Sudan and even Ethiopia, with concentrations in northern Nigeria and northern Cameroon. Another widely spoken Atlantic language is Wolof, the major indigenous language of Senegal.

The Kru languages are spoken in Liberia and southwestern Côte d'Ivoire, with relatively small numbers of speakers. Kru was included in Kwa (see below) by Greenberg (1963).

The Gur (Voltaic) group cover most of Burkina Faso, spreading also into northern parts of countries to the south. The Gur language with by far the largest number of speakers is Moore, the dominant indigenous language of Burkina Faso. One language sometimes considered to be Gur is Dogon, spoken around Bandiagara in Mali and adjacent parts of Burkina Faso, but current opinion questions this assignment and in Bendor-Samuel (1989) Dogon is considered at least provisionally a separate branch within Volta-Congo.

To the south of the Gur languages and continuing to the coast are the Kwa languages, stretching roughly from the Bandama River in the west to the Benin–Nigeria border in the east. The precise extent of Kwa has shifted considerably since Greenberg (1963), and not all the innovations have been generally accepted. In Bendor-Samuel (1989), the term Kwa covers essentially Greenberg's Western Kwa, with his Eastern Kwa being mostly reassigned to Benue-Congo (see below). The least controversial part of these changes is the exclusion of Kru (see above) and Ijo (see below) from Kwa. In what follows, as in the geographical description given above, the restricted sense of Kwa as in Bendor-Samuel (1989) will be followed. Kwa languages, in this narrow sense, include Baule, an important regional language of southern Côte d'Ivoire; the Akan dialect cluster (Twi-Fante), the major indigenous language of Ghana; the Ga-Dangme dialect cluster, including Ga, the major indigenous language of the Ghanaian capital Accra; and the Gbe dialect cluster, including Ewe, a widely spoken indigenous language in Ghana and Togo, and Fongbe, the most widely spoken indigenous language of Benin.

Ijo, now usually considered a distinct branch of Niger-Congo, is spoken around the delta of the Niger River in Nigeria, and is the major indigenous language of Nigeria's Rivers State. Different varieties of Ijo are not all mutually intelligible, the most prestigious varieties being Kolokuma and Kalabari.

The Adamawa-Ubangi languages are spoken in a belt from eastern Nigeria into Sudan, with the main concentration in the Central African Republic. The languages of the Adamawa subgroup are spoken to the west, those of the Ubangi subgroup to the east. The most widely used Adamawa-Ubangi language is Sango, which is the national language of the Central African Republic; historically, it is a creole derived primarily from the Ubangi language Ngbandi.

The remaining branch of Niger-Congo, Benue-Congo, covers most of sub-Saharan Africa from the western border of Nigeria eastward to the Indian Ocean and southward to the Cape. Most of this area and population falls under Bantu, but from a historical linguistic viewpoint Bantu is a rather low-level subgroup within Benue-Congo and the present geographical distribution of Bantu is the result of an expansion from the Nigeria–Cameroon border area that took place for the most part within the last two millennia. The most widely spoken Benue-Congo languages outside Bantu are Yoruba, an official language in southwestern Nigeria; Edo, to the southeast of Yoruba; Nupe, to the northeast of Yoruba; Igbo, an official language in central southern Nigeria; Ibibio-Efik to the east of the Niger delta in Nigeria; and Tiv, a regionally important language of eastern Nigeria.

As already implied, the Bantu languages occupy most of Africa from the Nigeria–Cameroon border to the east and south, including several major indigenous languages. The most widely spoken Bantu language is Swahili, originally the language of Zanzibar and the adjacent coast, although it has now spread as a lingua franca and also, especially in Tanzania, as a first language across large parts of east Africa; it is the official language of Tanzania and an official language in Kenya. Comorian, the indigenous language of the Comoros, is closely related to Swahili. Several other widely spoken Bantu languages are here listed primarily by country: Fang (Equatorial Guinea, Gabon), Bangala (Congo–Kinshasa), Kituba (Congo–Kinshasa), Lingala (Congo–Kinshasa), Kikongo (Congo–Kinshasa, Angola), Luba-Kasai (Congo–Kinshasa), Luba-Shaba (Congo–Kinshasa), Zande (Congo–Kinshasa and neighboring countries), Northern Mbundu (Angola), Southern Mbundu (Angola), Gikuyu (Kenya), Kamba (Kenya), Luyia (Kenya), Luganda (Uganda), Nyankore (Uganda), Soga (Uganda), Kirundi (Burundi), Kinyarwanda (Rwanda) – Kirundi and Kinyarwanda are essentially different national variants of the same language – Chagga (Tanzania), Haya (Tanzania), Makonde (Tanzania, Mozambique), Nyamwezi (Tanzania), Sukuma (Tanzania), Lomwe (Mozambique), Makua (Mozambique), Sena (Mozambique), Tsonga (Mozambique, South Africa), Nyanja (Malawi, Mozambique, Zambia), Tumbuka (Malawi, Zambia), Yao (Malawi, Tanzania), Nyakyusa-Ngonde (Malawi, Tanzania), Bemba (Zambia), Luvale (Zambia), Tonga (Zambia), Northern Ndebele (Zimbabwe), Shona (Zimbabwe, Zambia, Mozambique), Tswana (Botswana, South Africa), Southern Sotho (Lesotho, South Africa), Swati (Swaziland, South Africa), Northern Sotho (Pedi) (South Africa), Tsonga (South Africa), Venda (South Africa), Xhosa (South Africa), Zulu (South Africa).

4.3 Nilo-Saharan families

Nilo-Saharan, as proposed by Greenberg (1963), has proven to be more controversial than either Afroasiatic or Niger-Congo, although the most recent survey of the Nilo-Saharan languages (Bender 1997) is positive. The internal structure of Nilo-Saharan is also more controversial. In what follows I have therefore limited myself to citing some of the more widely spoken Nilo-Saharan languages and the branches of the family to which they belong.

Nilo-Saharan languages are not spoken in a continuous geographical area, and even in the areas mentioned below they are often interspersed with Afroasiatic (Chadic, Cushitic, also Arabic) and Niger-Congo languages. One Nilo-Saharan area is the middle course of the Niger River; another is Chad; a third is the Nile around the Egypt–Sudan border; while a fourth includes parts of southern Sudan, westernmost Ethiopia and Eritrea, northeastern Congo–Kinshasa, and parts of Kenya and Uganda to the north and east of Lake Victoria.

The westernmost language, or rather cluster of closely related languages, assigned to Nilo-Saharan is Songay, spoken along the Niger river in an area including the town of Timbuktu, although it is also the living language whose inclusion in Nilo-Saharan has proven most controversial (Bender 1997: 59). Another major western Nilo-Saharan language, assigned to the Saharan branch of the family, is Kanuri, the dominant indigenous language of Bornu State in northeastern Nigeria. Within the For(an) branch, mention should be made of For (Fur), spoken in the Darfur region in west-central Sudan.

Most of the more widely spoken Nilo-Saharan languages belong to the East Sudanic and Central Sudanic branches. East Sudanic includes the Nubian languages of the Egypt–Sudan border area, of which the most widely spoken is Nobiin. It also includes the Nilotic languages, a grouping which includes the Luo (Lwo) languages Acholi (Uganda), Lango (Uganda), Alur (Uganda, Congo–Kinshasa), and Luo (Dholuo) (Kenya); the Dinka-Nuer languages Jieng (Dinka) (Sudan) and Naadh (Nuer) (Sudan); the Eastern Nilotic languages Maasai (Kenya, Tanzania), Turkana (Kenya), Karamojong (Uganda), and Teso (Uganda, Kenya), and the Southern Nilotic language Kalenjin (Kenya). Central Sudanic includes Ngambay (Sara-Ngambay) (Chad), Lugbara (High Lugbara) (Congo–Kinshasa, Uganda), Mangbetu (Congo–Kinshasa), Ndo (Congo–Kinshasa), and Badha (Lendu) (Congo–Kinshasa).

It has been suggested that Meroitic, the extinct language of the Meroë civilization (ca. 2300–1600 BP), might be a Nilo-Saharan language, but Bender (1997: 32) considers the available data insufficient to resolve the issue. Finally, it should be noted that there are some as yet virtually undescribed languages spoken in the general Nilo-Saharan area that are as yet insufficiently known to establish whether or not they might be Nilo-Saharan. Many Nilo-Saharan and possible Nilo-Saharan languages are spoken in regions of current unrest (southern Sudan) or recent unrest (Ethiopia), which accounts in part for the rather poor state of our knowledge of such languages.

4.4 Khoisan families

The Khoisan languages are spoken predominantly in southwestern Africa. The area occupied by Khoisan languages has certainly contracted as a result of the spread of Bantu and, more recently, Indo-European languages, and all Khoisan languages have small numbers of speakers, with the largest, Nama (Khoekhoe), spoken in Namibia and South Africa, having an estimated 146,000 (Grimes 1996a: 323). Two otherwise unclassified languages of East Africa, namely Hadza and Sandawe of Tanzania, were proposed for inclusion in Khoisan by Greenberg (1963). However, even the genetic unity of Khoisan with the exclusion of Sandawe and Hadza is not accepted by all specialists, some of whom prefer to treat Northern Khoisan, Central Khoisan, and Southern Khoisan as distinct families. Sands (1998) is a recent treatment, concluding that there are striking parallels among the three nuclear branches of Khoisan plus Hadza and Sandawe, but that it is not clear to what extent this reflects common genetic origin versus contact.

4.5 Proposals for larger groupings

The proposal that Afroasiatic might form part of a larger Nostratic macrofamily was discussed in Section 2.7. Proposals that Niger-Congo and Nilo-Saharan might be distantly related are considered worth following up both in Bendor-Samuel (1989: 8–9) and by Bender (1997: 9).

5 Languages of the Americas

The internal and external genetic affiliations of the indigenous languages of the Americas have given rise to considerable debate in recent years, with proposals ranging from a total of three families (Greenberg 1987) to almost 200 (Campbell 1997). Since Campbell (1997) lists securely assured genetic units, but then also discusses proposals for broader genetic groupings, his account can serve as a survey that covers the range of proposals. In the space available, it would not make sense to list and discuss up to 200 genetic units, so in what follows a very selective choice will be made, concentrating on larger families and languages with larger numbers of speakers.

5.1 Languages of North America

The languages of North America are surveyed in Mithun (1999).

The northern fringe of North America is home to the Eskimo-Aleut family. The family has two branches, Aleut, spoken on the Aleutian Islands, and Eskimo. The latter starts in eastern Siberia and then stretches from Alaska to Greenland. Eskimo is properly a family of languages, with a major division between Yupik (Siberian and southern Alaskan varieties) and Inuit-Inupiak in northern Alaska, Canada, and Greenland. Greenlandic is the variety with most speakers, and is the national language of Greenland.

Another major family of North America is Na-Dene, although the precise extent of the family is controversial. Its core is Athabaskan, comprising most of the languages of the interior of Alaska, northwest Canada, with some languages (all extinct or moribund) in Oregon and northern California, and then a flowering in the geographically remote Apachean languages of the southwestern USA, including Navajo. It is established that the recently extinct Eyak language, spoken at the mouth of the Copper River in Alaska, is genetically related to Athabaskan, to give Athabaskan-Eyak. Less certain is whether Tlingit (spoken on the Alaska panhandle) is related to these, which would justify the more inclusive term Na-Dene, and even more questionable whether Haida (spoken on Queen Charlotte Island) should be added.

Other language families of the Pacific Northwest include Wakashan (British Columbia and adjacent Washington state; the family includes Nootka) and Salishan (British Columbia, Washington state, with some excursion into Idaho and Montana). Other language families of California, sometimes extending to adjacent areas, are Miwok-Costanoan, Chumashan, and Yuman. The small Keresan family and the language isolate Zuni of New Mexico, though small in number of speakers (each in the thousands) are among the most vigorous indigenous languages of the USA, with high rates of acquisition by children.

The Siouan languages are a major language family of the North American Plains, stretching from north of the US–Canada border through the Dakotas, Minnesota, and Wisconsin down to Arkansas, with outliers historically almost as far south as the Gulf and in Virginia. The Muskogean family, formerly concentrated in the southeastern USA, includes Choctaw, Chickasaw (these two arguably dialects of a single language), Alabama, and Seminole.

The Iroquoian languages are spoken around the Great Lakes, apart from Cherokee, originally spoken in Georgia; the family also includes Tuscarora, Huron (extinct), Seneca, and Mohawk. The Algic (Algonquian-Ritwan) family covers much of the northeast of North America, though also extending into western Canada and with two outliers on the Great Plains (Cheyenne and Arapaho). The family includes Blackfoot, the various forms of Cree spoken in Canada, and Ojibwa in Canada and the USA. These are all Algonquian languages. The two Ritwan languages, Wiyot (extinct) and Yurok (moribund), though indubitably related to Algonquian, are spoken in California.

Uto-Aztecan is one of the major language families of North America, spreading also into Meso-America. The Northern Uto-Aztecan languages include Shoshone, Comanche, Ute, and Hopi; while the Southern Uto-Aztecan languages include Pima-Papago (O'odham) in Arizona and Sonora; Cora and Huichol in Nayarit and Jalisco; and Nahuatl in central Mexico. Nahuatl was the language of the Aztec Empire.

5.2 Languages of Meso-America

The languages of Meso-America are surveyed in Suárez (1983). The Uto-Aztecan family was discussed in Section 5.1.

Other major language families of Meso-America are Otomanguean, Mixe-Zoquean, and Mayan. The Otomanguean languages are spoken mainly across the isthmus of Mexico, especially its southern part (Guerrero, Oaxaca, Puebla), including the Zapotecan languages and Mixtec, although some Otomanguean languages, such as Otomí, are spoken further north and

separated from the mass of Otomanguean languages. Mixe-Zoquean languages are spoken in a number of geographically separated groups in the isthmus of Mexico; the Olmecs, the first of the great Meso-American civilizations, seem to have spoken a Mixe-Zoquean language. The Mayan languages cover or covered most of Mexico east of the isthmus and also Guatemala and Belize; individual languages include Yucatec, Chol, Kekchi, and K'iche' (Quiché); although Chol does not have one of the highest numbers of speakers among Mayan languages, it is important historically as the most direct descendant of the language of the Mayan hieroglyphic inscriptions.

For Chibchan languages of Meso-America, see Section 5.3.

5.3 Languages of South America

For Amazonian languages, Dixon and Aikhenvald (1999) provide a survey; a comparable survey for Andean languages by Willem Adelaar and Pieter Muysken (2004) is available. Campbell's (1997) discussion of South American languages is based on Kaufman (1990).

The Chibchan language family includes a number of languages scattered from Costa Rica in the west through Panama to Colombia in the east. Cariban languages are scattered across northeastern South America, mostly to the north of the Amazon, although some languages are spoken as far west as Colombia and there is a geographically isolated group well to the south along the upper course of the Xingu River.

The precise extent of the Arawakan family is a matter of ongoing debate, although the group of languages that are clearly genetically related are sometimes referred to as Maipurean. They are scattered at great distances from one another across much of northern South America, from the Caribbean coast as far south as Paraguay, with Garífuna (also misleadingly known as Black Carib) spoken in Central America.

Tucanoan languages are spoken in northwestern South America (Colombia, Ecuador, Peru, adjacent parts of Brazil). Panoan languages are spoken in the Peru–Brazil border area, with some spillover into Bolivia, while the Tacanan languages, now believed to be related to Panoan in a Panoan-Tacanan family, are spoken in Bolivia with some spread into Peru. The Gê (Je) family is spoken in Brazil. Tupian languages are spoken both on the Amazon River and its tributaries and in an area that includes Paraguay and adjacent parts of Brazil, Bolivia, and Argentina. The most widely spoken Tupian language is Paraguayan Guaraní, spoken by 95 percent of the population of Paraguay and a national language of the country.

Quechumaran includes the Quechua and Aymara branches, although the nature of the relationship between Quechua and Aymara – genetic or contact – continues to be debated. Quechua is strictly speaking a language family rather than a single language, since different varieties are not mutually intelligible; in terms of numbers of speakers, it is the largest indigenous language family of the Americas. Quechua was the language of the Inca Empire, and partly as a result of this empire and later use as a lingua franca by the Spanish administration it achieved a spread from Colombia in the north to Argentina in the south, although most speakers are in Peru. The most widely spoken Quechua languages are South Bolivian Quechua in Bolivia, Cuzco Quechua in Peru, and Chimborazo Quichua in Ecuador. Aymara is spoken predominantly in western Bolivia.

5.4 Proposals for larger groupings

For proposals that would group Eskimo-Aleut as part of Eurasiatic, see Section 2.7. For proposals that would group Na-Dene with Sino-Tibetan and possibly other families, see Section 3.10. Otherwise, the main proposal is that of Greenberg (1987) to group all the remaining indigenous languages of the Americas into a single Amerind family.

There are also more modest proposals for larger genetic units within the indigenous languages of the Americas, excluding Eskimo-Aleut and Na-Dene, of which Penutian and Hokan

are perhaps the most engrained in the literature. Penutian would group together a number of languages and language families of the northern part of western North America, including at least: Maidu and Miwok-Costanoan (together California Penutian); Chinookan; Coos, Kalapuyan, and Yakonan (together Oregon Penutian); Klamath-Modoc and Sahaptin (together Plateau Penutian); and Tsimshian. Hokan would group together a number of languages and language families of the southern part of western North America and extending into Meso-America, including at least: Yuman; Karok-Shasta, Pomo, and Yana (Northern Hokan); Chumash, Salinan, and Seri; Tequistlatecan (Chontal of Oaxaca); and Washo.

6 Pidgin and Creole Languages

Since the main concern of this survey is the geographic distribution of languages as spoken by native speakers, pidgin languages (Holm 1989) will only be considered to the extent to which they are being creolized. Pidgin languages that are relevant in this way include the closely related Krio of Sierra Leone, Pidgin of Cameroon, and Pidgin of Nigeria, all of which are English-based pidgins undergoing creolization and widely used as lingua francas in the relevant countries. In addition, mention must be made of the closely related Tok Pisin of Papua New Guinea, Pijin of the Solomon Islands, and Bislama of Vanuatu, all likewise English-based pidgins undergoing creolization and widely used as lingua francas.

Otherwise, creole languages are particularly prevalent in the Caribbean and the islands of the Indian Ocean. They include English-based Sranan, the lingua franca of Suriname, and the French-based creoles of Haiti in the Caribbean and Mauritius, Réunion, and the Seychelles in the Indian Ocean. The case of Sango, in origin a creolized form of Ngbandi, was discussed in Section 4.2.

7 Deaf Sign Languages

Most works on languages of the world deal exclusively with spoken languages, and this is certainly the emphasis of this survey. Recent work on deaf sign languages has shown, however, that deaf sign languages are languages in their own right, differing considerably in structure from the spoken languages used in the same territory (see Chapter 18). Indeed, genetic relations among deaf sign languages often do not match those of the "corresponding" spoken languages, e.g. American Sign Language (ASL) is more closely related to French Sign Language than it is to British Sign Language. Lewis, Simon, and Fennig (2016) list 141 deaf sign languages, though without giving any internal genetic classification, and it is unfortunately true to say that our knowledge of all but a handful of deaf sign languages (such as ASL) is so poor that it is not at present possible to undertake such a task. This is clearly an area that merits further investigation.

REFERENCES

Abondolo, D. (ed.) (1998). *The Uralic Languages*. London: Routledge.

Adelaar, W. F. H. and Muysken, P. C. (2004). The Languages of the Andes. Cambridge Language Surveys. Cambridge: Cambridge University Press.

Ball, M. (1993). *Phonetics for Speech Pathology*, 2nd edn. London: Whurr.

Bender, M. L. (1997). *The Nilo-Saharan Languages: A Comparative Essay*, 2nd edn. Munich: Lincom Europa.

Bendor-Samuel, J. (ed.) (1989). *The Niger-Congo Languages*. Lanham, MD: University Press of America.

Benedict, P. K. (1975). *Austro-Thai: Language and Culture*. New Haven, CT: HRAF Press.

Blust, R. A. (2013). *The Austronesian Languages*, rev. edn. Canberra: Australian National University Research School of Pacific and Asian Studies.
Campbell, L. (1997). *American Indian Linguistics: The Linguistic History of Native America*. New York: Oxford University Press.
Casad, E. H. (1974). *Dialect Intelligibility Testing*. Norman, OK: Summer Institute of Linguistics.
Comrie, B. and Corbett, G. (eds) (1993). *The Slavonic Languages*. London: Routledge.
Diakonoff, I. M. (1988). *Afrasian Languages*. Moscow: Nauka.
Dixon, R. M. W. (1980). *The Languages of Australia*. Cambridge: Cambridge University Press.
Dixon, R. M. W. (1997). *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
Dixon, R. M. W. and Aikhenvald, A. (1999). *The Amazonian Languages*. Cambridge: Cambridge University Press.
Foley, W. A. (1986). *The Papuan Languages of New Guinea*. Cambridge: Cambridge University Press.
Greenberg, J. H. (1963). *The Languages of Africa*. Bloomington, IN: Indiana University and The Hague: Mouton.
Greenberg, J. H. (1971). The Indo-Pacific hypothesis. In T. E. Sebeok (ed.), *Linguistics in Oceania: Current Trends in Linguistics*, Vol. 8 (pp. 807–71). The Hague: Mouton.
Greenberg, J. H. (1987). *Language in the Americas*. Stanford, CA: Stanford University Press.
Harris, M. and Vincent, N. (eds) (1988). *The Romance Languages*. London: Routledge.
Hetzron, R. (ed.) (1997). *The Semitic Languages*. London: Routledge.
Holm, J. A. (1989). *Pidgins and Creoles*, 2 vols. Cambridge: Cambridge University Press.
Illič-Svityč, V. M. (1971–84). *Opyt sravnenija nostraticeskix jazykov*, 3 vols. Moscow: Nauka.
Johanson, L. and Csató, E. (eds) (1998). *The Turkic Languages*. London: Routledge.
Kaufman, T. (1990). Language history in South America: what we know and how to know more. In D. L. Payne (ed.), *Amazonian Linguistics: Studies in Lowland South American Languages* (pp. 13–67). Austin, TX: University of Texas Press.
Konig, E. and van der Auwera, J. (eds) (1994). *The Germanic Languages*. London: Routledge.
Lewis, F. P., Simons, G. F., and Fennig, C. D. (2016). *Ethnologue: Languages of the World. 19th edn*. Dallas, TX: Summer Institute of Linguistics. Also available online at https://www.ethnologue.com
Loprieno, A. (1995). *Ancient Egyptian*. Cambridge: Cambridge University Press.
MacAulay, D. (ed.) (1993). *Celtic Languages*. Cambridge: Cambridge University Press.
Masica, C. P. (1991). *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.
Matisoff, J. (2002). *Languages of Mainland Southeast Asia*. Cambridge Language Surveys. Cambridge: Cambridge University Press.
McAlpin, D. W. (1981). *Proto-Elamo-Dravidian: The Evidence and its Implications*. Philadelphia, PA: American Philosophical Society.
Mithun, M. (1999). *The Languages of Native North America*. Cambridge: Cambridge University Press.
Nettle, D. (1999). *Linguistic Diversity*. Oxford: Oxford University Press.
Nichols, J. (1992). *Linguistic Diversity in Space and Time*. Chicago, IL: University of Chicago Press.
Nikolayev, S. L. and Starostin, S. A. (1994). *A North Caucasian Etymological Dictionary*. Moscow: Asterisk.
Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.
Pawley, A. (1998). The Trans New-Guinea phylum: a reassessment. In R. A. C. Dam, C. Odé, and J. Miedema (eds), *Perspectives on the Bird's Head of Irian Jaya, Indonesia* (pp. 655-90). Amsterdam: Rodopi.
Posner, R. (1996). *Romance Languages*. Cambridge: Cambridge University Press.
Ramat, A. Giacalone and Ramat, P. (eds) (1998). *The Indo-European Languages*. London:
Routledge.
Ruhlen, M. (1987). *A Guide to the World's Languages*, Vol. 1. *Classification*. Stanford, CA: Stanford University Press.
Ruhlen, M. (1994). *On the Origin of Languages: Studies in Linguistic Taxonomy*. Stanford, CA: Stanford University Press.
Sands, B. (1998). *Eastern and Southern African Khoisan: Evaluating Claims of Distant Linguistic Relationship*. Cologne: Koppe.
Shibatani, M. (1990). *The Languages of Japan*. Cambridge: Cambridge University Press.
Sohn, H. (1999). *Korean*. Cambridge: Cambridge University Press.
Steever, S. B. (ed.) (1998). *The Dravidian Languages*. London: Routledge.
Suárez, J. A. (1983). *The Mesoamerican Indian Languages*. Cambridge: Cambridge University Press.
van Driem, G. (1997). Sino-Bodic. *Bulletin of the School of Oriental and African Studies*, 60, 455–88.
Wurm, S. A. (1982). *Papuan Languages of Oceania*. Tübingen: Narr.

3 Typology and Universals

WILLIAM CROFT

1 Introduction: The Typological and Generative Approaches to Language Universals

Typology represents an approach to the study of linguistic structure that differs in certain important respects from the generative and the functionalist approaches (see Chapters 7 and 8), although it is closer in spirit to the latter. The most important difference between typology and these other approaches to linguistic structure is that the typological approach is fundamentally crosslinguistic in nature. A generative linguist can analyze a single language in the search for universals of language structure. There are generative analyses of many different individual languages, but relatively few crosslinguistic generative studies (a notable exception is the work of Mark Baker, e.g., Baker 2003; but see Croft 2008). A functionalist linguist can also analyze a single language in the search for universals of the relationship of language structure to language function. Some functionalist theories are more extensively supported by crosslinguistic data, notably Functional Grammar (Dik 1997) and Role and Reference Grammar (Van Valin and LaPolla 1997). These theories recognize the variation in grammatical structures across languages, but their frameworks generally focus on what is believed to be common to all languages (see Section 5).

Typology, on the other hand, is fundamentally comparative. A genuinely typological analysis of a grammatical construction, or a phonological pattern, or other aspects of language, examines the variation across a large number of languages. In this respect, typology resembles comparative historical linguistics. The goals of typology and comparative historical linguistics are very different, however, although the results of each are essential to the other (see Section 2). Comparative historical linguistics seeks genetic (family tree) relationships among languages, in order to discover the history of the languages and their speech communities (see Chapter 15). Typology examines a broad sample of languages in order to discover universals of language structure and propose explanations for those language universals.

For this reason, typology is linked to language *universals*. For some linguists, typology simply means the description of variation, that is, how languages differ in their structure. For example, a simple descriptive typology of the word order of numeral and noun would divide the languages of the world into three broad *types*:

- those in which numerals normally precede the noun they modify, as in English (Indo-European) *two women*;

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

- those in which numerals normally follow the noun they modify, as in Ma'di (Nilo-Saharan, Uganda-Sudan) *àgɔ̄ sū* 'men four';
- those in which numerals may either precede or follow the noun they modify, as in Wardaman (Australian, Australia) *guyaminyi mulurruwuyayi* 'two old_women' or *marluga lege* 'old_man one.'

However, typology in the linguistic sense is more than a classification of how languages differ in their structure. A descriptive typology leads to generalizations that constrain how much languages can vary; those generalizations are language universals. Several examples of language universals will be given in this chapter.

The search for language universals is shared by typology and generative grammar. However, the language universals derived from typological research are quite different from those derived in generative grammar. Although the belief in language universals has considerable modern currency, it is by no means a necessary fact or universally held opinion, and in fact the opposite view was widely held in American linguistics until around 1960. To a considerable degree, the difference between the generative and typological approaches to language universals can be traced to the different traditions to which their respective founders, Noam Chomsky and Joseph Greenberg, responded. The generative approach represents a reaction against behavioristic psychology, while the typological approach represents a reaction against anthropological relativism.

The behaviorist view of language, in particular language learning, is anti-universalist in that it posits no innate, universal internal mental abilities or schemas. In the behaviorist view, linguistic competence is acquired through learning of stimulus-response patterns. In contrast, the generative approach posits the existence of innate internal linguistic abilities and constraints that play a major role in the acquisition of language. It is these constraints that represent language universals in this approach. The argument used by Chomsky (e.g., Chomsky 1976) for the existence of innate universal linguistic competence refers to the "poverty of the stimulus." It is argued that the child has an extremely limited input stimulus, that is, the utterances that it is exposed to from the mother and other caregivers. This stimulus is incapable of permitting the child to construct the grammar of the adult's language in a classic behaviorist model; therefore, the child must bring innate universals of grammatical competence to bear on language acquisition. Hence the primary focus on universals in the generative tradition has been on their innate character.

The anthropological relativist view of language is that the languages of the world can vary arbitrarily: "languages could differ from each other without limit and in unpredictable ways," in a famous quotation from the linguist Martin Joos (Joos 1957: 96). This view of language was particularly strong among anthropological linguists studying North American Indian languages, which indeed differ radically in many ways from so-called Standard Average European languages. However, the comparison of one "exotic" language or a limited number of languages to English only indicates diversity, not the range of variation, let alone limits thereto. Greenberg and others discovered that a more systematic sampling of a substantial number of languages reveals not only the range of variation but constraints on that variation. Those constraints demonstrate that languages do not vary infinitely, and the constraints represent language universals. Hence the primary focus on universals in the typological tradition has been on their crosslinguistic validity, and on universals that restrict language variation (see Section 5).

The innate universals posited by generative grammar are intended to explain linguistic structure. The poverty of the stimulus argument is essentially a deductive argument from first principles (although it does make assumptions about the nature of the empirical input, and what counts as relevant input). The poverty of the stimulus argument is one aspect of Chomsky's more generally *rationalist* approach to language. The universals posited by typology are intended to represent inductive generalizations across languages, in keeping with typology's *empiricist* approach to language. The generative grammarian argues that the discovery of innate principles that the child brings to bear in learning a single language can be extrapolated to language in general (Chomsky 1981). The typologist argues that a grammatical analysis based on one language

or a small number of languages will not suffice to reveal language universals; only a systematic empirical survey can do so.

The typological approach to discovering language universals, like any empirical scientific approach, is basically *inductive*. In order to discover what language in general is like, or the universal character of language, one should look at a large number of languages. For example, consider the two English examples in (1)–(2):

(1) They talked about the war.
(2) What did they talk about?

The question in (2) differs from the statement in (1) in three ways: the interrogative pronoun corresponding to *the war* in (1) is found at the beginning of the question; the preposition *about* nevertheless remains "stranded" in its position after the verb; and the auxiliary verb *did* occurs, positioned before the subject. Looking only at English, one cannot tell what the significance of each of these three differences is. In fact, the initial position of the interrogative pronoun is quite common in the world's languages, though by no means universal. In contrast, the stranding of the preposition and the insertion of an auxiliary that is absent in the corresponding statement are extremely rare and idiosyncratic grammatical traits of English, and not causally connected to each other or to the initial position of the interrogative pronoun. It is only by examining a broad sample of languages that the significance of different properties of linguistic structure can be assessed.

Likewise, one can only judge whether a causal connection exists between two grammatical properties by examining a large and widely distributed sample of languages. For example, languages differ in the constructions used for nonverbal predication. English requires a *copula* (in boldface in (3)) for nonverbal predication at all times, whereas Russian lacks a copula in the present tense:

(3) She **is** a doctor.
(4) Ona vrač.
    She doctor

A common explanation for the absence of a copula, or *zero* copula, in a variety of theories is that the reason for the occurrence of a copula is a "need" to place verbal inflections that cannot occur on nonverbal predicates such as the nouns in (3)–(4) (see Stassen 1997: 66 for the history of this proposal, and also Baker 2003: 40). Zero copulas therefore are claimed to occur when verbs lack inflection or when the inflection is zero; nonzero inflections require a "dummy" copula to carry them. Thus the zero copula occurs in Russian present tense because that is the "unmarked" inflectional category.

This hypothesis appears to be a plausible one, assuming that the purpose of the copula is to carry verbal inflections for nonverbal predicates. It appears to be valid for English, Russian, and a number of other languages. However, Stassen demonstrates that in fact it is not crosslinguistically valid in general (Stassen 1997: 65–76). There are many languages with overt inflections but a zero copula, such as Sinhalese (Indo-European, Sri Lanka; see p. xvii for abbreviations of grammatical elements in the examples):

(5) mahattea e-nəw-a
    gentleman come-NPST-IND
    'The boss comes/will come.'

(6) mahattea a-aw-a
    gentleman come-PST-IND
    'The boss came.'

(7) unnæhee hungak prəsiddə kene-k
3SG.M very famous person-NOM
'He is/was a very famous person.'

And there are many languages with (at least some) zero inflections but always an overt copula (translated as COP in example (9)), such as Cambodian (Austroasiatic, Cambodia):

(8) vìːə tʏu phsaːr
he go market
'He goes/went/will go to market.'

(9) mən-s nùh cìːə kruː
man that COP teacher
'That man is a teacher.'

Instead, the universal governing occurrence of a copula is based on the semantic type of predication: if predicate adjectives (predication of a property) require a copula in a language, then predicate nominals (predication of object class) require a copula as well (Croft 1991: 130; Stassen 1997: 127).

2 How Many Languages Are Needed for a Typological Study?

It is impossible to examine all of the approximately six thousand languages of the world (see Chapter 2) in order to describe their typological variation and infer language universals. As with other sciences, only a sample of languages can be taken, and the universals that are inferred are dependent on the quality of the sample. A broad sample is necessary to achieve two primary goals: to reveal the diversity of linguistic structures (see Section 1), and to discover statistically valid causal connections between linguistic traits. These two goals are generally served by a variety sample and a probability sample respectively.

A variety sample seeks to maximize the likelihood of capturing all the diversity of the construction in question (e.g., numeral-noun constructions). The standard approach to variety sampling in typology is to ensure that the languages in the sample are drawn from as many different language families and different geographical regions as possible given the size of the sample (anything from a dozen to several hundred languages or more). The idea behind this approach is that one is more likely to find structural diversity among languages that are not historically related, either by common ancestry or by contact. German and Navajo are more likely to display greater structural diversity than German and Dutch, because German and Dutch are very likely to be similar due to recent common ancestry and to continued contact.

There is another source of typological diversity, however. Language change is a gradual process, and if a construction is undergoing change in a family, then different members of that family will often display rare "intermediate" types. For example, grammatical voice systems have been undergoing change in the Salishan and Austronesian families, and a survey of languages in those families reveals many rare types. So a strict variety sample should be supplemented by surveys of families in which the construction being examined is changing.

A probability sample is intended to capture the relevant properties of the population as a whole, that is, all the languages of the world, including quantitative correlations of structural traits that may indicate a causal connection between those traits. Ideally, this is achieved by a random sample of languages whose relevant traits are independent. In practice this ideal must be compromised. First, a sample will not be completely random because not all languages are

documented and documentation is nonrandomly distributed across languages. As more languages are documented, this problem becomes less severe. Second, one cannot guarantee that all cases are independent, since the traits may be historically connected. For example, the fact that the numeral precedes the noun in both English and German is probably due to their descent from a common ancestor language (Proto-Germanic), hence including both languages in a typology of word order is to be avoided. The likelihood that shared linguistic traits are historically related varies with trait: some linguistic structures are unstable and change frequently, so even a large sample is likely to contain independent cases.

With a large sample, the likelihood increases that the typological traits of two or more languages in the sample are historically related, either by descent from a common ancestor or through language contact, recent or ancient. In fact, many linguists believe that all modern languages may be descended from a single common ancestor, even if the time depth and amount of change from that time means we can never construct a complete language family tree with a high degree of confidence. Thus it is possible that some typological traits have been a result of the form of Proto-World, or at least some very ancient protolanguage. For example, the indigenous languages of Papua New Guinea that do not belong to the Austronesian language family are almost entirely subject-object-verb (SOV) in word order. The island of Papua New Guinea was colonized at least 35,000 years ago, if not earlier, and it is likely that the SOV word order of all the languages on Papua New Guinea is historically related through common ancestry or contact for at least 35,000 years.

From a traditional statistical point of view, the fact that even distantly related languages may share typological traits through historical relations is problematic. The best solution to the problem of historical relatedness is to use the linguistic family tree directly to model the evolution of typological traits through time and identify any causal relations between those traits. However, there is great controversy over families of a time depth greater than 5,000–8,000 years, and even among accepted families, their subgrouping is also quite uncertain. Thus, our ability to use the language family tree to model causal relations in the evolution of language traits is limited at present.

Another approach is precisely to take historical relatedness into consideration when constructing and evaluating language universals. For example, Dryer (1989) does not discard any language from his sample, even if closely related genetically or in the same geographical area. Instead, he aggregates the data from closely related languages into *genera* (a genetic grouping of approximately the same time depth as a major subgroup of Indo-European such as Romance or Germanic), and then aggregates those results into six large linguistic areas in the world (Africa, Eurasia, Southeast Asia and Oceania, Australia and Papua New Guinea, North America, and South America). For example, using this method, Dryer demonstrates that the hypothesized relationship between adjective-noun order and object-verb order is a consequence of a high preponderance of this correlation among genera in the Eurasian area, and is not an inherent causal connection between these two word orders (Dryer 1988).

More recently, Maslova and her colleagues have developed a method of taking pairs of closely related languages, inferring from the match or mismatch of typological traits in those pairs the likely rate of change of the trait in a mathematical model of language change, in order to determine what the stationary distribution of the typological traits are, and hence their inherent dominance or preference (Maslova 2000). For example, in an unpublished paper Maslova and Rakhilina propose that the current distribution of alignment patterns between ergative and accusative (see Section 4) is not a stationary distribution: it appears that a stationary distribution would imply a higher proportion of the accusative alignment pattern, and hence a higher dominance of the accusative pattern.

It is also possible to use mixed effect models to consider the effect of language family membership and geographical proximity (and hence language contact) on the distribution of typological traits, as shown in Jaeger et al. (2011). A drawback is that a very large sample size is required

to obtain statistically significant results for the independent effects of language family membership, geographical proximity, and genuine causal connections between typological traits.

The integration of historical relations between languages in inferring language universals requires a shift to analyzing language structure as part of a dynamic evolving linguistic system over time, and language universals as universal constraints on the evolution of linguistic systems. This shift in approach, necessitated by methodological issues in constructing language samples, actually dovetails with the general shift in typological theory from universals of synchronic language systems to universals constraining language change (Section 6).

3 How Does One Person Use Data from So Many Languages?

A question that often arises about typological research, especially outside the field of linguistics, is how a single typologist or a small group of collaborators can acquire the data from the large number of languages in a sample? It is impossible for a single linguist to know directly a hundred or a few hundred languages, or to consult with a similar number of native speaker consultants. For that reason, typological research is necessarily built on the foundations of sound documentation of the world's languages by native speakers or by field linguists and their native speaker consultants (see Chapter 4). Many typologists have also done fieldwork in particular languages, language groups, or language areas. The severe endangerment of the vast majority of languages in the world has long been of great concern to typologists.

Typologists have used a variety of resources to obtain the data for typological research. Perhaps the most important is a reference grammar of a language. A surprisingly large percentage of the languages of the world have reference grammars, though they vary substantially in breadth and depth of coverage and in quality. A reference grammar is quite different from a pedagogical grammar: it is designed not for learning a language, but as a reference source for the major grammatical constructions of the language. A reference grammar is organized around families of constructions (and also the sound structure and morphological structure of words in the language), although this organization varies from grammar to grammar. A reference grammar presupposes to a greater or lesser extent a knowledge of basic linguistic terminology and a knowledge of the language itself (for example, many reference grammars give examples only in the script of the language).

In order to use data from a language that one does not speak or has not studied in detail, a linguist must be able to analyze the structure and the meaning of phrases or sentences in the language. The meaning is generally provided by a translation of an example into English (or French or Spanish, or whatever language the reference grammar is written in). The structure of the example is ideally provided in a separate line, sometimes referred to as the *interlinear morpheme translation* (IMT) by the author of the reference grammar. Examples of IMTs can be seen in examples (4)–(9) above. An example of a sentence from Mokilese, an Austronesian language of Micronesia, with an IMT is given in (10).

(10) mine woaroa-n woal-o war
exist CLF-3SG man-that canoe
'That man has a canoe.'

The last line, the *gloss*, gives the meaning of the Mokilese original sentence as a whole: from the gloss, we can see that this is an example of the predication of possession, to use the technical linguistic description of the function of the sentence. The middle line, the IMT, glosses each morpheme of the original, including affixes and other meaningful internal word changes, following relatively widely used conventions. The IMT provides the structural analysis of the sentence: the order of elements (e.g., the verb comes first in the sentence); which elements are independent

words and which are not (e.g., 'that' is a suffix on 'man'); and the presence of special grammatical elements (e.g., the possessive classifier for the man's canoe, and the person-number inflectional suffix on the classifier). The IMT also shows how differently the function of predication of possession is expressed in the language: in this case, a more literal translation of the Mokilese sentence would be 'That man's canoe exists.'

The presence of IMTs allows a linguist to rapidly recognize the grammatical structure of the language in the description even if the linguist has not previously studied the language. IMTs are used widely in contemporary reference grammars. Many grammars, especially older ones, do not use IMTs, and therefore a linguist must study the reference grammar (and a dictionary if available) in greater detail in order to analyze the structure of the constructions of interest in the language.

Reference grammars vary significantly in the constructions that they cover, and the depth in which they cover them. A typologist may discover that some (or many) of the reference grammars do not describe the construction s/he is interested in. One may need to accept that there will be gaps in the data in the sample (or one will replace the language in the sample with another language possessing a grammar with the relevant description). More recently, typologists have developed questionnaires to obtain data for particular constructions which are typically poorly described in reference grammars. One of the earliest questionnaires used a set of 200 sentence contexts to elicit tense-aspect constructions; the questionnaire was distributed to native speakers or specialists in 65 different languages (Dahl 1985). More recently, experimental elicitation techniques have been used to avoid the problem in questionnaires of translation from the language of the questionnaire. For example, a set of 71 pictures of different spatial scenes was used to elicit constructions for spatial relations in nine diverse languages (Levinson, Meira, and the Language and Cognition Group 2003), and a set of 61 video clips of different cutting and breaking scenes was used to elicit different verbs of cutting and breaking in 28 languages (Majid et al. 2007).

Questionnaires and experimental elicitation techniques have the advantage of allowing a linguist to directly elicit the constructions expressing the function s/he is investigating. However, designing questionnaires and elicitation strategies is extremely difficult. Also, questionnaires and elicitation, while valuable in themselves, can benefit from a broader description of the language than just the constructions under investigation. Many times a typologist discovers an unexpected correlation of constructions. For example, one might not consider the English conditional construction in (11) and the topic construction in (12) to be grammatically related:

(11) If he comes, I will stay.

(12) (As for) me, I will stay.

But in a number of languages, including Hua, a Papuan language of New Guinea, the antecedent of a conditional and a topic are marked grammatically in the same way (in the case of Hua, with the suffix *-ve;* Haiman 1978: 570–1):

(13) e-si-*ve* baigu-e
come-3SG.FUT-*ve* will.stay-1sg
'If he comes, I will stay.'

(14) dgai-mo-*ve* baigu-e
I.EMPH-CONN-*ve* will.stay-1sg
'(As for) me, I will stay.'

This grammatical connection is at first surprising, although an explanation is forthcoming: the situation described by the antecedent of a conditional provides the framework for evaluating

the consequent that follows it, and the referent described by a topic phrase constitutes the framework for the following discourse (Haiman 1978: 577–86). This grammatical and semantic connection would probably not have been detected if one constructed a questionnaire for conditional constructions, but it might be observed if one consulted reference grammars for a wide range of languages and noticed the similarity of form. Thus reference grammars continue to have great value for typological research.

4 How Can One Compare Grammatical Structures from Many Different Languages?

A fundamental characteristic of the typological approach to language is that one begins by comparing a wide range of languages for the grammatical construction in question. The most important observation that comes from that process is the extraordinary diversity of grammatical structures that are found. For example, we noted above that English expresses the predication of possession with a transitive verb *have* as in *That man has a canoe*, but the Mokilese expression is literally "That man's canoe exists," with an intransitive verb "exist." Other constructions for the predication of possession would translate literally as "A canoe is located at that man," "That man is with a canoe," and "As for that man, a canoe exists" (Heine 1997: 92). Each of these constructions has its own distinctive grammatical structure, and in addition to these major types there are a number of other less common types.

What is the basis for crosslinguistic comparison? The tremendous diversity of grammatical structures points to an answer that was offered in the seminal paper in typology, Greenberg (1966), and has been reiterated since: meaning or function. For example, in the study of possession, Heine compares whatever grammatical constructions are used in a language to express the function of predication of possession. Possession is what is traditionally described as a semantic relation: an ownership relation between a possessor (usually a person) and a possessum (usually an artifact). Predication is generally described as a pragmatic function, having to do with the packaging of information in a sentence: the possession relation is asserted, in contrast to attributive possession (as in *that man's canoe*), in which the possessive relation is used to modify the description of the referent *canoe*.

The use of function as a basis for crosslinguistic comparison allows one to avoid problems arising from the usage of traditional grammatical terminology. For example, in English a contrast is made between relative clauses with *finite* (fully inflected) verb forms, as in (15), and *participles*, a nonfinite verb form, as in (16):

(15) the papaya that the ants ate
(16) the papaya eaten by the ants

A crosslinguistic study of "relative clauses" quickly reveals that some languages such as Turkish (Altaic, Turkey) use participial or nominalized verb forms heavily, as in example (17) (Comrie 1989: 142; see Keenan and Comrie 1977):

(17) Hasan-ın Sinan-a ver-diğ-i patates-i yedim
Hanan-GEN Sinan-DAT give-NR-his potato-ACC I.ate
'I ate the potato that Hasan gave to Sinan.'

However, this is an artifact of the traditional grammatical terminology, which treats finite and nonfinite constructions completely differently. Keenan and Comrie define relative clauses functionally, in terms of a proposition modifying a referent (the noun); using this definition, Turkish has constructions comparable to English relative clauses (and participles).

One widely held but invalid assumption about the crosslinguistic comparison of grammatical structures is that one can only compare grammatical structures by using the same grammatical categories across languages. Languages can vary to a remarkable degree in basic categories of grammar. For example, a plausible candidate for a pair of universal grammatical categories are the categories of subject and object of a verb:

(18) The woman didn't run.

(19) The snake bit the man.

The sentence in (18) has only a single phrase (*the woman*) referring to a participant in the event denoted by the verb (*run*). Such a sentence is intransitive, and *the woman* is labeled the "subject"; we will refer to the intransitive "subject" with the label S. The transitive sentence in (19) on the other hand has two phrases referring to the two participants in the event (*bit*). It seems completely natural, indeed even necessary, that the first phrase, *the snake* (labeled A, mnemonic for "agent") should belong to the same grammatical category as *the woman* in (18). Both *the woman* and *the snake* occur before the verb. Substitution of a pronoun for *the woman* would require the subject form *she*, not *her*. The grammatical category grouping S and A is traditionally called "subject." The second phrase in (19), *the man* (labeled P, mnemonic for "patient") is grammatically different. It occurs after the verb, and substitution of a pronoun for *the man* in (19) would require the object form *him*, not *he*. The grammatical category consisting of P is generally called the "object."

But many languages do not categorize the phrases referring to the participants in events in the same way. Compare the translations of (18) and (19) in Yuwaalaraay, an Aboriginal language of Australia:

(20) wa:l ṇama yinar-Ø banaga-ṇi
not that woman-ABS run-NFUT
'The woman [S] didn't run.'

(21) ḍuyu-gu ṇama ḍayn-Ø yi:-y
snake-ERG that man-ABS bit-NFUT
'The snake [A] bit the man [P].'

Yuwaalaraay does not have subject and object in the English sense. The grammar of participants is expressed by case suffixes on the noun. In an intransitive sentence like (20), the phrase labeled S has no suffix (notated here with the zero symbol -Ø). In a transitive sentence like (21) however, what an English speaker would call the "subject," A, has a case suffix *-gu*, which is called the *ergative* case (abbreviated ERG), and the "object" phrase P has no suffix, like the "subject" S in (20). In other words, whereas English categorizes both A and S together (as subject) and distinguishes P (as object), Yuwaalaraay categorizes P and S together (called the *absolutive*, abbreviated ABS) and distinguishes A (as the ergative).

This difference between English (and many other languages) on the one hand, and Yuwaalaraay (and many other languages) on the other, is very striking. It seems very unnatural to us to group together S and P against A – subject and object in the English sense seem to be such basic categories of grammar. Since the terms "subject" and "object" are so loaded with theoretical and traditional connotations, most typologists discard them in describing the contrast between languages like English and languages like Yuwaalaraay, and instead use the terms *nominative* and *accusative*, from the Latin case terms, for the English categories. The crosslinguistic difference between English and Yuwaalaraay is called a difference in *alignment*.

The difference between English and Yuwaalaraay grammatical patterns here seems to make the two languages entirely different, so that no comparison, let alone universals of language, can be derived from the study of alignment systems. But this is not true. Crosslinguistic comparability

of grammatical relations is based on the semantic roles, or more precisely prototype clusters of semantic roles, represented by S, A, and P: one can compare how the agent of "bite" and the single participant of "run" are encoded grammatically across languages. More strikingly, there are language universals that can be stated no matter what the alignment system of the language is. For example, the absolutive is zero coded in Yuwaalaraay and many other languages, in contrast to the overt coding of the ergative case. In languages with a nominative-accusative alignment and case inflections, the nominative is usually zero coded in contrast to overt coding for the accusative. These are not the only types found: the universals for each alignment system can be more precisely formulated to cover all types, as in (22a–b):

(22) a. The accusative is encoded with at least as many morphemes as the nominative.
b. The ergative is encoded with at least as many morphemes as the absolutive.

The universals for the two alignment systems in (22a–b) have a single explanation. Noun phrases with nominative referents are more frequent in discourse than noun phrases with accusative referents, and noun phrases with absolutive referents are more frequent in discourse than noun phrases with ergative referents. A far-reaching universal encompassing (22a–b) and many other grammatical categories is: categories of meanings that are more frequent in discourse are likely to be encoded with fewer morphemes (Croft 2003; see Section 6). Hence it does not matter what the grammatical categories of a language are in order to find language universals. One must simply identify how concepts are encoded in a language. The language universals are based on how concepts are encoded, not a set of universal grammatical categories.

The methodological necessity of comparing constructions encoding the same functions across languages matches the functionalist orientation of most typologists. In order to discover language universals, a typologist compares how function is encoded in linguistic form. A crosslinguistic perspective does not allow a typologist to analyze form autonomously from its function. The very fact of crosslinguistic variation in the formal encoding of linguistic function, however, implies that linguistic form is separate from linguistic function, and that linguistic form is at least partly arbitrary – otherwise, all languages would have the same grammatical structure.

5 The Nature of Language Universals

There is a widely assumed view of language universals that language universals are simply properties that are possessed by all human languages. All linguists would agree that there are certain very basic properties that are possessed by all languages. These properties include: utterances made up of discrete meaningful units; conventions of syntax; and duality of patterning (that is, organization of sound structure at a level largely independent of the organization of syntactic structure). These language universals are generally described as *design features* of language, and are examples of *unrestricted* language universals.

The more controversial question is over the existence and nature of universals that make reference to more specific grammatical entities. Some proposed universals of this type would include the hypothesis that all languages make a distinction between nouns and verbs, or that all languages have consonants and vowels.

Typological research leads to two general conclusions about language universals that are more specific reference to grammatical categories and structures. The first is that language universals are typically not in the form, "All languages have X": that is, language universals are not unrestricted universals. Instead, almost all of the language universals that have been discovered are restricted or *implicational* universals, in the form "If a language has X, then it also has Y."

Consider, for example, the relative orders of certain types of modifiers, in particular adjectives and numerals. In Section 1 we noted that in some languages, the numeral precedes the noun it

modifies and, in others, it follows (and in still others, either order is possible). The same is true of the order of an adjective and the noun it modifies. When both modifiers are compared in a single language, the picture changes. In English both adjectives and numerals precede the noun:

(23) a. red book b. three books
Adj Noun Num Noun

This pattern is found in many languages. In many other languages, both adjectives and numerals follow the noun:

(24) Kosraean (Autronesian, Caroline Islands)
a. mwet kuh b. mwet luo
men strong men two
Noun Adj Noun Num

A third group of languages has adjectives following the noun while numerals precede:

(25) Jamiltepec Mixtec (Mixtecan, Mexico)
a. vēhē lúhlu b. uvi vēhē
house little two house(s)
Noun Adj Num Noun

On the other hand, languages with the adjective preceding and numeral following are virtually unattested (although there are a few; the existence of exceptions will be discussed below).

The pattern of attested vs. unattested (or rare) language types can be given in the four-cell Table 3.1, with numbers of languages in the World Atlas of Language Structures (WALS) Online.

The generalization can itself be described in terms of an *implicational universal*:

(26) If a language has adjective-noun word order, then it (almost always) has numeral-noun word order.

The discovery of implicational universals of word order by Greenberg (1966) demonstrated that there exist universal properties of language that do not require all languages to be identical in specific properties. The implicational universal in (26) is not by itself a description of a fact about the grammar of a particular language. In fact, one could not even identify the implicational universal without looking across a set of languages. The implicational universal captures a contingent relationship between adjective-noun order and numeral-noun order. Nevertheless, this contingent relationship must be a part of individual speakers' knowledge of language structure and meaning. In particular, a language will not (or is extremely unlikely to) emerge or change to a type that has both noun-numeral order and adjective-noun order.

Hundreds of implicational universals have already been discovered, and more are discovered every time a typologist investigates a new area of grammar. The existence of so many implicational universals requires a rethinking of the nature of Universal Grammar, which is usually thought of as a set of unrestricted universals. The part of Universal Grammar that consists of

Table 3.1 Adjective and numeral word orders, with numbers of languages on WALS Online (accessed 14 January 2011; 842 languages total).

	Noun-adjective order	Adjective-noun order
Numeral-noun order	attested (149)	attested (228)
Noun-numeral order	attested (433)	**very rare (32)**

unrestricted universals specifying ways in which all languages are identical captures only a very small portion of what is universal about language. It misses most of what is universal about language beyond the basic design features. The presence of large numbers of implicational universals requires a model which allows for a great deal of variation in grammatical structure across languages, but constrains that variation to a significant degree in many different dimensions. This basic observation about language universals has not been addressed by most syntactic theories.

The second general conclusion about language universals that is revealed by typological research is that the constraints on language variation that are represented by implicational universals are not exceptionless in the way that the design features of language are. In the example of the relationship between numeral-noun order and adjective-noun order given above, we noted that the fourth language type, noun-numeral order and adjective-noun order, is not actually unattested: a small number of languages of this type do exist. This is almost always observed for language universals, especially with the advent of larger and larger language samples.

This conclusion also changes how language universals must be understood. One cannot conceive of language universals (beyond the design features) as specifying what constitutes a possible human language. The noun-numeral and adjective-noun language type is not impossible; a few such languages exist. But it is far less frequent than the other three types. Moreover, the other language type with modifiers on the opposite side of the noun, adjective-noun and noun-numeral, while frequent, is not as frequent as the types in which both modifiers precede or both modifiers follow. These differences in the likelihood of language types are significant and must be explained by linguistic theory. In other words, typology shifts the scientific question about language universals from "What is a possible language type, and why?" to "What is a more probable language type, and why?"

For this reason, typologists have turned to quantitative methods in order to identify valid language universals. If all types of languages exist in the area of grammar under investigation, then one must be confident that differences in likelihood are statistically valid. The increase in quantitative sophistication in language sampling was discussed in Section 2. In the case of implicational universals, Maslova (2003) gives statistical tests for identifying valid implicational universals when one or more types in a table such as Table 3.1 is of much lower frequency than the other types.

An important type of model for inferring the language universals underlying generalizations such as the implicational relation in (26) is the *competing motivations* model. A competing motivations model posits two or more factors that determine language structure. However, the motivations typically do not determine a single grammatical pattern because they are often in conflict. In the case of conflict, there is no single optimal grammatical pattern that satisfies all of the competing motivations, and instead one finds crosslinguistic variation over several suboptimal patterns. In this way, universal properties of language (the motivations) give rise to crosslinguistic diversity.

For example, Greenberg proposed two competing motivations for implicational universals of word order. The first, *dominance*, can be thought of as simply a default preference for one order over another. For example, noun-adjective order (NA) is dominant, as is numeral-noun order (NumN). The second, *harmony*, can be thought of as a dependent relation of one word order upon another. For example, AN order is harmonic with NumN order and NA order is harmonic with NNum order.

Greenberg's two motivations compete with each other, and the result is described in the following principle:

(27) A dominant order may occur at any time, but a recessive order occurs only when a harmonic order is also present.

The principle in (27) accounts for the distribution of languages in Table 3.1. The upper left cell is the language type with both dominant orders (NA and NumN), which are not harmonic with each other. The other two attested types have one recessive order, but the harmonic order is also present. The extremely rare type would have both recessive orders (AN and NNum), neither of which is dominant. That is, the extremely rare type is not motivated by either dominance or harmony, which accounts for its rarity. Finally, one cannot satisfy both motivations at once, since the dominant orders are not harmonic with each other.

A competing motivations model can be found by inspection with a simple case such as the two interacting word orders in Table 3.1. But it turns out that comparing the word orders of several different constructions across languages is much more complex: most possible types are attested, but at highly varying frequencies. Justeson and Stephens (1990) use a statistical technique, log-linear analysis, to examine the relationships between multiple word orders. Log-linear analysis allows one to construct a model of the simultaneous interactions between many word orders. The model that best fits the data provides the best model of which word orders actually might be causally connected. The best fit model for the word orders investigated by Justeson and Stephens identifies the dominance and harmonic relations that are actually supported by the data in their 147-language sample. Reassuringly, Justeson and Stephens arrive at the same result that Dryer came to (see Section 2), namely that there is no direct causal relation between adjective-noun order and object-verb order.

Another example of the use of sophisticated statistical techniques in typology is found with an important recent development in modeling language universals, the *semantic map* model. The semantic map model allows the typologist to identify language universals without assuming that grammatical categories are the same across languages, as we observed in Section 4. There, it was pointed out that some languages divide the semantic roles in transitive and intransitive verbs (A, P, and S) in a nominative-accusative pattern (A+S vs. P) while others divide them in an ergative-absolutive pattern (A vs. S+P). Still other languages use a tripartite pattern, distinguishing all three of A, S, and P, while others have a neutral pattern, not distinguishing any of them. No language (or virtually no language) groups together A and S and distinguishes that group from P.

The range of variation and the limitation of attested types of grammatical relations can be represented by a *conceptual space* linking A to S and P to S, as in Figure 3.1. The attested language types can be mapped as connected regions on this three-point space. The unattested (or extremely rare) type of A+P vs. S cannot be mapped on the conceptual space in Figure 3.1: A cannot be grouped with P without also including S, according to the connections in Figure 3.1. Hence the conceptual space represents what is universal about the relationships among A, S, and P, while allowing for the variation of attested language types illustrated by the semantic maps in Figure 3.2.

The semantic map model is a powerful tool for identifying language universals and separating language universals from the arbitrary aspects of crosslinguistic variation, and has been widely used. However, it is impractical for identifying the conceptual space when that space consists of many more points than the three exemplified in Figures 3.1–3.2, or when the range and frequency distribution of attested language types is complex. However, these problems can

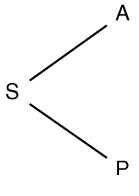


Figure 3.1 Conceptual space for semantic roles.

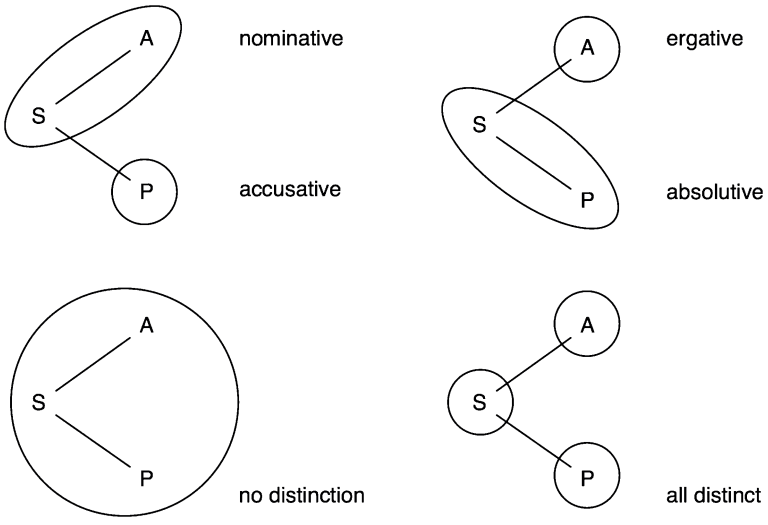


Figure 3.2 Map of attested systems of grammatical relations.

be solved by using multidimensional scaling and related multivariate techniques. For example, multidimensional scaling was applied to the results of the experimental elicitation for the 71 pictures of spatial relations described in Section 3, allowing typologists to identify the most significant semantic dimensions of spatial relations in determining attested systems of spatial relations (prepositions, postpositions, and case inflection; Levinson, Meira, and the Language and Cognition Group 2003; Majid et al. 2007; Croft and Poole 2008; Croft 2010).

The richness, diversity, and complexity of typological evidence are daunting to analyze. Typology has only recently begun to use more sophisticated quantitative tools to allow researchers to go beyond the easy to identify language universals to the universals that are harder to detect through the noise of arbitrary crosslinguistic variation. The initial successes in using such tools indicates that typologists will be able to find many more restricted universals of language.

6 Explanations for Language Universals

Most typologists seek functional explanations for language universals. This is chiefly because the basis of crosslinguistic comparison is how function is encoded in grammatical form. Nevertheless, there are many aspects of language function, and different aspects have been proposed to explain different kinds of language universals.

Proposed explanations for universals of word order (and also the order of affixes) fall into two general categories: language processing in production and comprehension, and diachronic explanations. Most models of word order universals include harmonic patterns, in which two or more word orders are correlated. The processing explanation is based on theories of how easy or difficult it is for speakers construct or parse utterances: put crudely, the word order combinations that facilitate parsing are more common crosslinguistically, and those that make parsing more difficult are less common (a detailed theory is presented in Hawkins (1994, 2004). A diachronic explanation of word order correlations is based on the fact that the actual grammatical constructions used for different word orders are historically related. For example, there is a very tight harmonic correlation between genitive-noun order and the order of adposition and noun:

genitive-noun correlates with postpositions, and noun-genitive with prepositions (Greenberg 1966). The most likely explanation for this correlation is that adposition constructions are very frequently historically derived from genitive constructions via the process of grammaticalization (Greenberg 1969; Lehmann 1995; Bybee, Perkins, and Pagliuca 1994; Hopper and Traugott 2003). This can be observed in the history of English, where the prepositional phrase *inside the house* derives historically from the noun-genitive construction *in the side of the house*. Grammaticalization theory has uncovered a large number of universals of language change, specifically processes by which words in particular syntactic constructions evolve into grammatical elements.

The semantic map model also lends itself to a functional explanation for patterns of the expression of grammatical relations (see Figure 3.2 in Section 5), spatial relations, and many other grammatical categories and constructions. The explanation is based on a universal conceptual space. The conceptual space represents the degree of conceptual relatedness of the situations or concepts represented by the points in the conceptual space. The structure of the conceptual space is hypothesized to represent general properties of human cognition and conceptualization. The language-specific categories represented by the semantic maps are partly arbitrary (hence the crosslinguistic variation), but they must conform to the constraints imposed by the structure of the conceptual space (Croft and Poole 2008).

Another widely used functional explanation appeals to competing motivations of economy and iconicity (Haiman 1983, 1985). A simple example of language universals where economy and iconicity are involved is the expression of inflectional categories. In the category of nominal number, many languages express the singular form without any inflection (zero coding), while the plural is expressed with an overt inflection, as in English *branch ~ branch-es*. Other languages express both singular and plural with overt inflection such as the Zulu (Bantu, South Africa) prefixes in (28):

(28) a. umu-ntu
sg-person
b. aba-ntu
pl-person

Other languages, such as Lahu (Sino-Tibetan, Burma) in (29), make no distinction, or, to put it another way, express both the concepts of singular and plural without any overt inflection:

(29) qhâʔ ‘village/villages’

However, very few languages express the plural without an overt inflection and the singular with an overt inflection. (In the case of languages that do, the plural is designated a collective and the singular is a special singulative form, and indeed this pattern is typically associated with nouns for objects occurring in groups.)

The typological pattern can again be described in terms of a table (Table 3.2) and an implicational universal.

(30) If a language uses an overt inflection for the singular, then it also uses an overt inflection for the plural.

The variation allowed by the implicational universal in (30) can be accounted for by economy and iconicity. *Economic* motivation is characterized in processing terms: the more frequently

Table 3.2 Attested and unattested singular and plural inflectional types.

	Overt plural inflection	No plural inflection
No singular inflection	attested (English)	attested (Lahu)
Overt singular inflection	attested (Zulu)	**extremely rare**

occurring form is expressed by fewer morphemes. The singular is more frequent than the plural in discourse for the vast majority of noun referents, so it may be expressed without an inflection, as in English. *Iconic* motivation is characterized by the relationship between form and function: it favors a one-to-one mapping between form and function. So if the category of number is expressed in a language, iconicity motivates expression of all values of the category, as in Zulu. (Alternatively, the category may not be expressed at all, as in Lahu.)

Economic motivation is extremely common crosslinguistically: many universals of grammatical inflection are explainable in terms of frequency differences, and economy is increasingly being invoked to explain universals of syntactic constructions as well (Bybee 2006). Iconic motivation is also extremely common. In fact, iconicity is often taken for granted. For example, it is simply assumed that referents and their modifying properties form a syntactic constituent (e.g., in *Two boys ate five pizzas* the number of boys is two and the number of pizzas is five, not the other way around); but it represents the iconic motivation that conceptual relations support syntactic constituency. Also, most syntactic theories contain principles specifying that the syntactic arguments of a verb must match the semantic participants in the event denoted by the verb in number and type; this is also an iconic principle. Economy and iconicity are frequently in competition, leading to the crosslinguistic variation of the type observed in Table 3.2.

The competition between processing mechanisms of different kinds and the different types of conceptual relations holding among components of the situation expressed by a grammatical construction leads naturally to a dynamic model of language and language universals. Many typologists have proposed evolutionary models of how language structures adapt to the functions they perform in communication and the constraints on comprehension and production of utterances in language use (e.g., Greenberg 1979; Croft 2000, 2003; Givón 2002). In an evolutionary approach, a linguistic system is conceptualized as a stage in the process of language change, which can shift by the choices of speakers. Language universals are the result of universal cognitive and social interactional forces that shape speakers' choices at all timescales, from the immediate discourse situation to the lifetime of a speaker to the transmission of language across generations. The evolutionary approach to grammar allows linguists to integrate crosslinguistic variation with variation in the social use of language (sociolinguistics; see Chapters 24 and 25) and in language use (exemplar and usage-based models; Bybee 2007, 2010). The variation found across languages in typological analysis forms the basis for the understanding of language structure and variability in a single speech community and in the mind and behavior of a single speaker.

REFERENCES

Baker, M. (2003). *Lexical Categories: Verbs, Nouns and Adjectives*. Cambridge: Cambridge University Press.

Bybee, J. L. (2006). From usage to grammar: the mind's response to repetition. *Language*, 82, 711–33.

Bybee, J. L. (2007). *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.

Bybee, J. L. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, J. L., Perkins, R. D., and Pagliuca, W. (1994). *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago, IL: University of Chicago Press.

Chomsky, N. (1976). *Reflections on Language*. London: Fontana.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Comrie, B. (1989). *Language Universals and Linguistic Typology*, 2nd edn. Chicago, IL: University of Chicago Press.

Croft, W. (1991). *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. Chicago, IL: University of Chicago Press.

Croft, W. (2000). *Explaining Language Change: An Evolutionary Approach*. Harlow: Longman.

Croft, W. (2003). *Typology and Universals*, 2nd edn. Cambridge: Cambridge University Press.

Croft, W. (2008). Methods for finding language universals in syntax. In S. Scalise, E. Magni, and A. Bisetto (eds), *Universals of Language Today* (pp. 145–64). Berlin: Springer.

Croft, W. (2010). Relativity, linguistic variation and language universals. *CogniTextes* 4, 303. Available online at http://cognitextes.revues.org/303/.

Croft, W. and Poole, K. T. (2008). Inferring universals from grammatical variation: multidimensional scaling for typological analysis. *Theoretical Linguistics*, 34, 1–37.

Dahl, Ö. (1985). *Tense and Aspect Systems*. Oxford: Blackwell.

Dik, S. C. (1997). *The Theory of Functional Grammar*, edited by K. Hengeveld (2 vols). Berlin: Mouton de Gruyter.

Dryer, M. S. (1988). Object-verb order and adjective-noun order: dispelling a myth. *Lingua*, 74, 185–217.

Dryer, M. S. (1989). Large linguistic areas and language sampling. *Studies in Language*, 13, 257–92.

Givón, T. (2002). *Bio-linguistics: The Santa Barbara Lectures*. Amsterdam: John Benjamins.

Greenberg, J. H. (1966). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (ed.), *Universals of Grammar*, 2nd edn (pp. 73–113). Cambridge, MA: MIT Press. (Reprinted in Greenberg 1990, pp. 40–70.)

Greenberg, J. H. (1969). Some methods of dynamic comparison in linguistics. In J. Puhvel (ed.), *Substance and Structure of Language* (pp. 147–203). Berkeley and Los Angeles: University of California Press. (Reprinted in Greenberg 1990, pp. 71–118.)

Greenberg, J. H. (1979). Rethinking linguistics diachronically. *Language*, 55, 275–90. (Reprinted in Greenberg 1990, pp. 673–91.)

Greenberg, J. H. (1990). *On Language: Selected Writings of Joseph H. Greenberg*, edited by K. Denning and S. Kemmer. Stanford, CA: Stanford University Press.

Haiman, J. (1978). Conditionals are topics. *Language*, 54, 564–89.

Haiman, J. (1983). Iconic and economic motivation. *Language*, 59, 781–819.

Haiman, J. (1985). *Natural Syntax: Iconicity and Erosion*. Cambridge: Cambridge University Press.

Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Heine, B. (1997). *Possession: Cognitive Sources, Forces, and Grammaticalization*. Cambridge: Cambridge University Press.

Hopper, P. and Traugott, E. (2003). *Grammaticalization*, 2nd edn. Cambridge: Cambridge University Press.

Jaeger, T. F., Graff, P., Croft, W. and Pontillo, D. (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology*, 15, 281–320.

Joos, M. (ed.) (1957). *Readings in Linguistics I*. Chicago, IL: University of Chicago Press.

Justeson, J. S. and Stephens, L. D. (1990). Explanations for word order universals: a log-linear analysis. In W. Bahner, J. Schildt, and D. Viehweger (eds), *Proceedings of the XIV International Congress of Linguists, Vol. III* (pp. 2372–6). Berlin: Mouton de Gruyter.

Keenan, E. L. and Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8, 63–99.

Lehmann, C. (1995). *Thoughts on Grammaticalization: A Programmatic Sketch, Vol. I*, rev. edn. Munich: LINCOM Europa. (Originally published in 1982 as *Thoughts on Grammaticalization: A Programmatic Sketch, Vol. I*. Arbeiten des Kölner Universalien-Projekts, 48. Cologne: Institut für Sprachwissenschaft. Revised edition reprinted as *Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt*, 9. Erfurt: Seminar für Sprachwissenschaft der Universität.)

Levinson, S. C., Meira, S., and the Language and Cognition Group (2003). "Natural concepts" in the spatial topological domain – adpositional meanings in crosslinguistic perspective: an exercise in semantic typology. *Language*, 79, 485–516.

Majid, A., Bowerman, M., van Staden, M., and Boster, J. S. (2007). The semantic categories of cutting and breaking events: a crosslinguistic perspective. *Cognitive Linguistics*, 18, 133–52.

Maslova, E. (2000). A dynamic approach to the verification of distributional universals. *Linguistic Typology*, 4, 307–33.

Maslova, E. (2003). A case for implicational universals. *Linguistic Typology*, 7, 101–8.

Stassen, L. (1997). *Intransitive Predication*. Oxford: Oxford University Press.

Van Valin, R. D., Jr. and LaPolla, R. J. (1997). *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.

4 Field Linguistics

Gathering Language Data from Native Speakers

PAMELA MUNRO

1 What Is "Field Linguistics"?*

Unlike most of the other subfields of linguistics described in this book, field linguistics is not a theoretical discipline, and there is relatively little literature devoted to this area. Field linguistics, as I will use the term here, refers to the collection of primary data on the basic grammatical facts of a relatively little studied language from ordinary speakers in a relatively natural setting, and to the analysis and dissemination of such data. This type of data collection is usually called "fieldwork." Classic fieldwork is done in "the field," the area where speakers actually live (rather than in an artificial setting, such as a university classroom or office), or, even more classically, the area from which the speakers' ancestors originated.

Many types of linguistic endeavor share some of these features of field linguistics:

- armchair linguistics, where a native speaker linguist reflects on his or her own judgments (often confirmed by questioning other speakers) and analyzes these;
- psycholinguistics, where speakers produce responses to highly controlled stimuli;
- language acquisition studies, where children's language development is observed, often in a completely natural setting;
- sociolinguistics, where speakers' linguistic behavior is observed and correlated with facts about their backgrounds.

* I am grateful to a number of colleagues who sent me their answers to this question and others I consider here: Aaron Broadwell, the late Ken Hale, Jack B. Martin, Laura Martin, Russell Schuh, and Siri Tuttle. I have learned a lot about fieldwork from observing and talking to many other linguists over the years, including Leanne Hinton and the late Margaret Langdon. I must also thank all the wonderful native speakers without whose help I could not call myself a field worker, especially those I mention here and in the references: the late Betty Bland, the late Pollyanna Heath, Anita Lambey-Martinez, Felipe H. Lopez, Maurice Lopez, the late Robert Martin, Martha Martinez-Ciego, and Catherine Willmond.

As always this is for Allen and JP, and dear Alex.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

Most people would agree, however, that none of these domains constitute field linguistics.

Although field linguistics can be done anywhere, it is not normally based on introspection: linguists working on introspective data usually are not field linguists, even if their language is quite exotic. Thus, the languages on which field linguistics is done typically have few if any native speaker linguists, and one of the priorities of some field linguists is to train native speakers in the techniques of linguistic analysis.

A native speaker linguist might certainly use introspection to produce data to be analyzed for a basic description of his or her language, but introspective armchair linguistics is normally directed at puzzling out relatively obscure or at least higher level problems in languages whose grammar is already fairly well understood. Similarly, psycholinguistic studies conducted in the laboratory, acquisition studies based on observation of children in their homes and elsewhere, and sociolinguistic studies conducted in a community generally do not have the goal of producing basic grammatical description.

Studies like these can succeed precisely because basic description already exists. The primary goal of field linguistics is to produce descriptions of languages – often the first such descriptions. For this reason, what I am calling field linguistics is also called descriptive linguistics.

There are many techniques for collecting data and doing fieldwork (see Sections 2 and 3). But data collection is only the first step. The data collected must be analyzed (see Section 4) and, very importantly, disseminated. (Data, even analyzed data, that remains in someone's notebook or computer or tapes is of little value to anyone.) Any circulated data must be written in a system that is analytically consistent and maximally useful to the widest range of users.

Although there is not much literature describing field linguistics as a discipline, the amount of linguistic literature that results from field linguistics is huge. The type of literature or other production that results from the analysis of field data can vary considerably. Basic descriptions usually take the form of grammars (or articles on grammatical topics) or dictionaries. These works often serve as sources for reanalyzing of the data, perhaps from a different theoretical viewpoint. Novel data from field linguistics has provided numerous vitally important insights to mainstream theoretical linguistics over the years, and may also be important for other scholarship (Section 5). Many serious field linguists, however, feel a compulsion to make the results of their fieldwork available to the communities of speakers who use the language being analyzed (Section 6). Because of these efforts, some field linguists may regard their work as having more social value than many ivory tower enterprises, though possibly such feelings arise in part as a reaction against an idea that more theoretically oriented linguists hold those who collect primary data in low esteem. (Is this true?)

Fieldwork is addictive, at least for some people. The reason I do fieldwork is that I feel energized and my spirits lift on days when I get to work with speakers, and I cherish my relationships with the speakers I work with.

2 How Is "Field" Data Gathered?

2.1 Basic techniques of field linguistics

Linguists gather data directly from native speakers of the languages under investigation. There are several ways in which this is done.

Most often, particularly in the early stages of fieldwork, a linguist uses an intermediary language in order to ask for translations of words, phrases, and sentences in the "target" language. This question and answer translation process is called "elicitation."

Some field linguists frown on the process of direct elicitation and prefer to work entirely from more natural "volunteered" data. Most often, this involves recording from the speaker an extended narrative of some sort (a "text"), such as a retelling of a traditional story or a personal reminiscence.

The linguist then works carefully through the text with the speaker, obtaining not only a careful transcription and translation but following up on grammatical constructions and paradigms that arise in the text, in order to put together a full description of the grammar of the language in the text. A counterpart to such text analysis, where possible, is observing natural conversations between speakers. Even if the linguist cannot understand everything that is being said, he or she can take note of new words and grammatical structures that may appear only in discourse.

Both techniques have their pluses and minuses. Beginning the study of a completely unfamiliar language with simple words in isolation is a good way to become familiar with the language's sound system; hearing words only in complex contexts can make phonetic contrasts harder to distinguish than when those words are uttered in isolation.

But simple elicitation is never sufficient in itself. If the linguist makes up all utterances for translation or comments by the speaker, there is a significant possibility of creating unnatural or skewed data. For example, the speaker's translations may be influenced by the structure of the intermediary language, or, when the linguist grows confident enough to create new forms and sentences on his or her own, the speaker may be too polite to reject these. (Consequently, it is important for the linguist to ask a speaker to repeat back any linguist-created sentence – if the speaker cannot repeat it, it is unlikely to be fully acceptable – and to carefully mark in his or her notes any sentence that was not spontaneously produced by a speaker. If a construction never occurs in spontaneous speech, but is only accepted on the linguist's model, it is unlikely to be a standard feature of that speaker's usage.)

Elicitation and textual analysis are important complements to each other. One cannot assume every grammatical structure will show up in a text, so it is important to elicit missing structures directly. On the other hand, texts and conversational data often may reveal words and structures that never appear in sentence elicitation.

Serendipitous events can produce spontaneous types of language that are hard to elicit and that may never appear in texts. I had studied the Muskogean language Chickasaw for eight years and hundreds of hours before I began bringing my new baby Alex to visit my Chickasaw teacher, Catherine Willmond. One day, she took him on her lap and patted with his hand on the table in front of them, telling him,

(1) Pas pas pas aachi![1]
pas pas pas say

I had never heard this type of sentence before, but after further work I learned that it was a type of "expressive" construction used to describe noises that speakers feel are particularly appropriate for illustration presented to children. (Catherine's remark could be translated either 'He's going *pas pas pas* (making a slapping noise)!', or as a command addressed to Alex, 'Go *pas pas pas* (make a slapping noise)!'. The sentence was especially striking because outside of words used in this construction, such as the expressive syllable *pas*, Chickasaw has no words ending in *s* – and many other expressive syllables exhibit similar phonological peculiarities (Munro 1998). I have also learned that the presence of a baby is helpful for stimulating a speaker to produce diminutive forms of verbs, which in a number of languages may be used to show that a verb has a small or dear subject (somewhat like honorific forms in many Asian languages) (Munro 1988). Of course I am not suggesting that all linguists should bring babies into the field as a standard prop. But it is important to follow the speaker's reactions and train of thought, and to pursue new lines of inquiry that are suggested by things that happen during the field session.

Up till now I have not considered monolingual fieldwork, in which both the linguist and the speaker communicate only in the target language. Complete monolingual fieldwork is rather rare, since it requires enormous dedication by the linguist, if he or she is to really achieve a level of fluency such that he or she can discuss the speaker's subtle judgments entirely in the target language. However, many other forms of grammatical study can be conducted monolingually,

or partly monolingually. One considerable benefit of any such work is that it increases the native speaker's respect for the linguist!

In what follows I will assume that field linguists will engage in some direct bilingual elicitation, but that this will be combined with other types of investigation.

2.2 *Getting started with fieldwork*

2.2.1 The field methods class Many linguists' first experience with working with a native speaker comes in a linguistic field methods course in graduate school. In such a class the students meet with a speaker of an unfamiliar language and elicit forms which they transcribe and analyze. Eventually, the students learn enough to have a fairly good understanding of the grammar of the language.

Part of field methods class involves learning what might be called politeness or respect. In certain stages of a field methods class, students occasionally become so excited by the data that they forget that it is being provided by a real person, with a real person's needs and feelings. (I've had students turn to me in the middle of a class elicitation session and say, "Why did he say that?," referring to the speaker in the third person, as if he would not understand or be interested in hearing himself discussed!) The respect that is due to the native speaker who assists with a field methods class necessitates finding a suitable word to refer to that speaker. Traditionally, the speakers who provide data for linguists have been called "informants," a word that originally had at least a neutral sense. In the last few decades, however (beginning with Watergate, I think), the English word *informant* has become a euphemism for *informer*, and it has acquired all the negative connotations of that word in the minds of most nonacademics.[2] I see no reason to apply such a loaded, unpleasant word to the wonderful people who introduce me and my students to the joys of their languages, so I don't use the word *informant*, and I don't allow my students to do so in my hearing. Having to think of a substitute term is positive, since it forces the linguist – or field methods student – to evaluate his or her own relationship with the speaker. The normal term I use is "consultant," but often (particularly when the speaker is older) "teacher" is more appropriate. Many of the native speakers who work with me are coauthors of books or papers about their languages: in this case, "collaborator" is often the best term.

My own field methods classes follow a traditional model. The students are not told what the target language is until the first day of class, and after that they are asked not to read any literature on the language until they have figured out certain aspects of its grammar for themselves (about halfway through the first of two ten-week terms). I have them begin by eliciting nouns (since in most languages these can be pronounced in isolation more readily than other types of words); the class members discuss together first their initial phonetic transcriptions and then their first ideas about what the phonological system of the language is (what the phonemes or distinctive opposing speech sounds are, in other words). After the class members have worked out their own phonological analyses, we compare these to existing ones in the literature – if any exist – and attempt to work out a consensus. (I discuss in Section 4.2 below the question of how words in the language are to be spelled.)

I don't allow students to tape-record early class sessions in field methods. The reason for this prohibition is that no one initially is very good at recording data from a new language, however hard they try: only practice and analysis develop this skill. But if they know that a tape recording is available, many students are less motivated to work hard on transcription. In theory, having a tape of the session should mean that the student could work diligently on improving his or her transcription later. But a tape is never as good as being there with the speaker, when you can listen again, ask for repetitions, ask for slower or faster versions, or look at the speaker from different angles, so I don't want students to adopt this crutch at the beginning. (We often do record a sample tape of interesting words after a few sessions, and students are welcome to tape sessions after they have learned to transcribe well, as long as they ask the speaker's permission. It is wrong to tape-record anyone without asking permission.)

Words in isolation are fairly easy for anyone to elicit from a speaker, but problems can arise when moving on to what seem like simple sentences. If I ask someone, "How do you say, 'I'm going'?," that person may tell me the way to say "I'm going" in his language, but he or she may also say "You're going," responding not to the metalinguistic translation task but treating the request like a real-world event. Students need to learn early that context is very important, since if the speaker imagines a different context from the one they have in mind, the result may be unexpected or confusing. Similarly, speakers learn how to interpret the strange questions linguists ask, and generally become much more tolerant of funny sentences. After students acquire a small vocabulary and learn something about the grammar of the language, they make up their own words and sentences, asking the speaker to judge if they sound all right. (This is a difficult skill for both the student and the speaker. Speakers sometimes feel it would be impolite to criticize an understandable but ungrammatical utterance by the linguist, while linguists in love with their own theories may not listen hard enough to the way the speaker may say, "Yes, you can say that.")

In the second major assignment the students have to work out how subjects and objects of different persons and numbers are marked in different types of clauses with different types of verbs. This task can be quite easy for some languages, quite difficult for others.

In American Indian languages, which have been the target of most of my own fieldwork, and which I try to use as in every graduate field methods course I teach, it is very common to find both subject and object marked with affixes on the verb or verb phrase. Sometimes such marking is quite transparent, but often it is not.

Table 4.1 presents the different verbal agreement markers in the Arawakan language Garifuna (spoken in Belize, Honduras, and neighboring regions of Central America). Markers in the P column are prefixes; those in the I column are infixes; and the remaining markers are suffixes. Each set of markers is distinguished for seven person-number categories: first person singular, second person singular, third person singular masculine, third person singular feminine, first person plural, second person plural, and third person plural. Although there is a considerable amount of overlap among the sets, they are all distinct.[3]

The different sets of Garifuna person markers are used in different syntactic environments: a member of one of the eight sets is used to mark the subject and then, in certain constructions, a marker from a second set may be used to mark an object. Thus, for example, in a sentence like

(2) N-áfaru ba-dibu. 'I will hit you'

1s-hit aux-2s

Table 4.1 Verbal agreement markers in Garifuna.

	P	T	D	D+	N	L	I	S
1sg.	n-	-tina	-dina	-dina	-nina	-lina	-ná-	-na
2sg.	b-	-tibu	-dibu	-dibu	-nibu	-libu	-bú-	-bu
3m.	l-	-ti	-i	-li	-ni	-li	-ní-	-i
3f.	t-	-tu	-u	-run	-nu	-lu	-nú-	-u
1pl.	wa-	-tiwa	-diwa	-diwa	-niwa	-liwa	-wá-	-wa
2pl.	h-	-türü	-dürü	-dürü	-nürü	-lürü	-nû-	-rü
3pl.	ha-	-tiyan	-yan	-yan	-niyan	-liyan	-yán-	-yan

two affixes are used: a P prefix *n-* on the verb *áfaru* 'hit'[4] and a suffix *-dibu* on the future auxiliary *ba*. Speakers' usage is entirely consistent, but understanding it is a challenge for the analyst, and requires a fairly extensive amount of data, since both the particular syntactic construction and the semantics of the specific verb involved determine which markers will appear. Because of the partial overlap between sets, only full paradigms can determine which set of markers is used in a given construction. Thus, for example, in (2), the *-dibu* suffix could belong to either the D set or the D+ set of markers. Only with further data, such as

(3) N-áfaru ba-i. 'I will hit them' (pronounced as *Náfaru bei*)[5]
1sP-hit aux-3pD

can we securely identify the suffixes in both (2) and (3) as belonging to the D set. And so on.

I have taught three field methods classes using Garifuna as a target language. Students have difficulty accepting how complex the pronominal agreement pattern is – they are reluctant to believe that a system can be this complicated, and often prefer to assume that they may have misrecorded *-tibu* as *-dibu*, for instance, or to simply ignore troublesome pieces of data. The lesson here is to look for patterns and to accept that the data may be complex if that is the only consistent explanation.

An important class activity is analyzing a text from the speaker. After the text has been recorded on tape, students go through it individually, producing their own transcriptions of what they heard on the tape. Then we go through the text carefully with the speaker, as described earlier. Transcribing recorded texts like this in a language that one does not know well is extremely difficult. Although it is possible to produce a quick-and-dirty transcription of a recorded text by simply playing a bit of the text, asking the speaker to repeat what was on the tape, and writing this down, the effort of transcribing the text beforehand is worthwhile. Often speakers are mistaken about what was on the tape, or they may change an incomplete portion of the text to make it sound better out of context. Frequently, we'll produce more than one version of the text – a fully accurate transcription of the recording, containing pauses, hesitations, false starts, and so on, and the speaker's edited version, with everything said the way the speaker feels is best. Each of these has different linguistic uses.

From collecting data, students move on to more extended grammatical description and analysis, choosing individual topics on which to write substantive papers based on individual elicitation with the speaker. Many students' field methods papers are later revised for publication, or may even be developed into masters' theses or dissertations.

Increased student facility with and access to computation has changed field methods. In my current classes, we enter all our data into an online database. Word processing makes paper writing smoother, and the collected data can be searched in many ways (see Section 3.1.3). Still, there is no substitute for just sitting and staring at the data, as all serious analysts know.

2.2.2 Finding a speaker The field methods class teacher locates a speaker and makes all the arrangements for that speaker to show up for class and elicitation appointments: student linguists just need to come to class and use their brains. (Of course, this is an ideal situation: field methods consultants are people, not data machines, and they may get sick or develop other conflicts in the middle of the term, posing logistical problems for the teacher.) But real fieldwork requires the linguist to find a speaker to work with, which may be easier said than done.

One might assume that one would choose a language first, then find a speaker, and this is, of course, what many people do. But many other linguists who want to do fieldwork – but who, perhaps, are located in places where few exotic languages are spoken – happily choose to study any language that they can find a speaker of.

There are many ways to find a speaker. Personal contacts and serendipity are often very important. Because I know that every couple of years I will be teaching field methods, I keep up my

contacts in the Los Angeles American Indian and indigenous communities. Los Angeles has a very large Indian population (largely relocated from reservations by now discontinued Federal programs), but increasingly fewer speakers of Indian languages, and it might take me 50 or more phone calls or emails to find someone. All Indian languages of the United States are endangered, most critically, so I know that eventually there will come a time when linguists from my university will not be able to find speakers of more than a few American Indian languages in the city. But as long as immigration from Latin America continues, there will a steady stream of speakers of indigenous languages from Mexico and further south. Many of these languages have never been described. We are lucky to have been able to study five different Latin American languages in field methods classes.

Many linguists have a lot of trouble explaining their theoretical interests even to members of their families. Before you look for a speaker to work with, it's important to consider how you will tell that person about your work and goals. I normally tell a speaker that I am interested in learning his or her language, and in my case (since I'm such a terrific language junkie), this is completely true. Field methods class presents a problem, however. I usually try to explain to prospective consultants that students take the class because (in our department) it is a requirement, and that they want to learn the process of learning a language from a speaker rather than from a language class or from books or tapes. But this sounds a little cold, and it's not surprising that speakers have trouble believing that the students really may not be interested in their language for its own sake. I urge the students, therefore, to try to develop such an interest – to read about the people and their culture rather than just about the language, and to work as hard as possible on their pronunciation. These things help validate their interest to the speaker, and increase the speaker's trust. Doing these things, even if they start out as conscious behaviors designed to impress, increases the chances that the linguist will be successful and really will learn a lot about the language.

3 What to Ask a Speaker, and What a Speaker Says

Some people begin fieldwork on a language with a definite question or agenda in mind. Perhaps they are researching a particular syntactic construction or phonological feature crosslinguistically, or maybe they are looking for data to compare with that in a related language they know better. Having too much of an agenda or coming to the work with too many assumptions, however, can produce unexpected results.

One linguist I know had an ambitious plan for a crosslinguistic study of the potential ambiguity of sentences with quantifiers, such as *Two men carried four boxes* (did they have a total of eight boxes, or only four boxes between them?). He had shown native speakers of a variety of languages cute pictures of various configurations of men and different types of boxes, with interesting results. When he asked the late Pollyanna Heath to describe the pictures in her language, Maricopa, however, he encountered problems. In Maricopa (as in many American Indian languages), verbs for various activities are selected based on the shape of affected objects. Since some of the boxes in the pictures were round and some were oblong, different verbs had to be used, and it was impossible to translate the sentences simply.

I was reminded that I didn't know everything about how to do fieldwork myself while I was studying Creek, a Muskogean language related to Chickasaw, which at the time I already knew very well. I was primarily eliciting Creek words to compare phonologically with those in other Muskogean languages, but also idly trying to learn a little about Creek grammar. After I had been working on Creek this way for about a year, I happened to ask my consultant, Betty Bland, for the translation of an English sentence containing a plural noun. I was chagrined to learn that Creek has noun plurals – I had never checked to find how these worked in Creek, because Chickasaw nouns have no plural form, and I wrongly assumed that Creek would share this feature.

For these reasons, it is good to begin work on a new language by doing a general survey of as many features of basic grammar as possible: verb and noun inflection, simple tense and modality, questions, negatives, existentials, passives (if they exist), causatives, reflexives and reciprocals, adpositional expressions, adjectives, adverbs, imperatives, and so on. This procedure reduces the chance of embarrassing surprises and often pays dividends in the form of revealing areas where the grammar is particularly worth studying.

If the language has been studied already, it is certainly worthwhile to review existing descriptions. These can be used to help develop a plan for elicitation sessions, and may speed analysis. Of course, earlier descriptions may not be correct, or may prove to be based on a different dialect from that of the current speaker, so important facts from such works should always be rechecked. (This is not the only thing that should be rechecked, of course. The linguist's own data, particularly old data, should be rechecked and added to regularly. It's horribly embarrassing to find that a crucial word or beloved sentence elicited only once and cited frequently since then in fact turns out not to be replicable!)

As in a field methods class, it's normally best to begin the study of any new language with simple words in isolation in order to develop a feeling for the phonetics. Nouns are usually more simply inflected than verbs, so they are often good to start with. If a full sentence is too difficult to hear all at once, one can ask the speaker to say parts of it on their own.

Certain types of phrases, however, are dangerous to elicit out of context. I find that speakers of many languages are uncomfortable translating complex noun phrases on their own, and often translate "the blue house" as 'The house is blue'. To see how to say "the blue house," then, it is usually best to find out how this phrase appears in a sentence like "The blue house burned down." (Actually, the same comment can apply to certain types of sentences. A complete sentence that may seem very easy to understand to you may be interpreted completely differently by the speaker. It is often useful to ask when a particular utterance would be used.)

It is always wise to note many things about elicited data. Obviously, if a speaker says that a sentence made up by the linguist is bad, that is worth noting, but it's also important for the linguist to make sure that a sentence he makes up that the speaker approves can actually be repeated. A sentence that the speaker says "sounds okay," but which he can't repeat back, is certainly not a perfect sentence. Similarly, a construction which the speaker agrees to and repeats willingly, but which he never volunteers himself either in translation or in other uses, is an odd construction, and it's worthwhile for the linguist to try to figure out why this pattern is avoided in natural speech. Finally, if a given sentence is translated by the speaker only with great difficulty, that should be noted too.

I try to always write down any comments the speaker makes about data we discuss. Catherine Willmond, my Chickasaw teacher and collaborator, occasionally says, "That's the way white people say it." This is a surprising comment, since I am the only non-Indian I have ever encountered who can speak Chickasaw at all, and the sentences in question are often completely novel for me. But evidently such sentences share some (incorrect!) feature with the speech of nonfluent speakers. I haven't figured this out yet, but I diligently note this comment each time it's made, along with other cryptic remarks. The late Robert Martin, my first Mojave teacher, would sometimes explain the difference between two apparently synonymous sentences by saying that one meant "You're saying it" and the other meant "You're telling him." This is another one I haven't figured out yet. But maybe someday I will!

3.1 Working in the field

3.1.1 Fieldwork can be done anywhere I have made many field trips away from home to study languages. I spend an average of a week or ten days in Oklahoma (studying Chickasaw, and earlier also Choctaw and Creek-Seminole) every year (accompanied by Mrs. Willmond, who I also

see regularly in Los Angeles), and I have worked on Yavapai, Mojave, Chemehuevi, and Hopi in Arizona; several varieties of Zapotec in Oaxaca; Kawaiisu in California; and Garifuna in Belize; I've also been privileged (with other linguists as guides) to hear Choctaw in Mississippi, Apache in Arizona, and Alabama and Koasati in Texas – in each case on one or more trips away from home. But all the hours on all these trips put together would constitute only a small percentage of my total hours spent doing fieldwork on days when I spend the night in my own bed. Most of my field data has been gathered from native speakers with whom I met during classes at UCLA, in my office, or in their homes no more than a few hours' drive around Los Angeles.

Now, in a few cases the speakers in question were actually in their original ("aboriginal") locations, since I've studied a number of California Indian languages that are still spoken less than half a day's drive from where I live, including Cahuilla, Kawaiisu, Tübatulabal, Luiseño, and Diegueño. Most of the speakers I've worked with in the Los Angeles area, however, are people who live in Los Angeles, but who were born elsewhere. Most or all of my work with speakers of Zapotec (several languages, originally spoken in Oaxaca); Garifuna (originally spoken in Belize); Lakhota (originally spoken in South Dakota); Pima, Yavapai, and Maricopa (originally spoken in Arizona); Navajo (originally spoken in Arizona and New Mexico); Chickasaw, Cherokee, and Choctaw (originally spoken in Oklahoma); Crow (originally spoken in Montana); Yupik Eskimo (originally spoken in Alaska); K'iche' and Q'anjob'al (originally spoken in Guatemala); and Quichua (originally spoken in Ecuador) was done in Los Angeles.

For the most part, linguistic data gathered away from speakers' traditional homelands can be just as valid as linguistic data gathered in those homelands. But of course there are tradeoffs.

An important worry for many linguists contemplating working with a displaced speaker is whether that person still commands his or her language as well as someone with the support of a whole community. This is a valid concern – anyone can forget his or her language with no practice or stimulation. But any minority language speaker – as almost all speakers of American Indian languages are these days – is in danger of not using his or her language enough. Displaced speakers sometimes use their language more than people back on the reservation – it all depends on their personal situation and circumstances. It is certainly important to chat with prospective consultants about how and how much they use their language. And consultants may well bring different types of experience to different tasks. A field methods class, for instance, is primarily studying a single speaker's usage patterns, so it is not crucial that that speaker be a conservative follower of standard grammatical descriptions.

If the linguist contemplates writing the first description of the grammar of a language, it is important to work with more than one speaker, if possible, and to supplement any work with displaced speakers with work in the homeland community. Even when the bulk of the work is done with a displaced speaker (such as Catherine Willmond, my Chickasaw collaborator, who has lived in Los Angeles since 1959), briefer exposure to other speakers can serve as a useful check on and addition to the data (for example, I have worked with over 50 other Chickasaw speakers in Oklahoma, some of them for over 20 years).

One really important and gratifying aspect of working with displaced speakers is that one can share the fieldwork experience with a much larger group of students and others than could ever come along on overnight excursions. I teach a UCLA class on American indigenous linguistics in which Mrs. Willmond used to come to campus once a week and help students learn about her language (while they studied other aspects of these fascinating languages, and learned how to access data from a grammar). Although these students have the experience of asking for words and sentences and learning how to "spell" (i.e., transcribe) them, this is certainly not fieldwork. But the experience introduces Chickasaw to students who not only have never heard an American Indian language, but have never met an American Indian. Mrs. Willmond has now retired from this work, but the class has continued with Mrs. Martha Martinez-Ciego and other Garifuna elders, which has been just as successful. Sure, we can tell people about endangered languages – but meeting a speaker of such a

language and experiencing firsthand the beautiful structures threatened with loss makes the point dramatically.

Certain types of field linguistics, however, can only be done where there are concentrations of speakers (as many as possible) located as near as possible to where their ancestors lived. Traditional dialect surveys are done only with the most conservative of speakers (never with those transplanted thousands of miles from home to a new multicultural environment); to be most useful, these should be done in the field. Other types of sociolinguistic data, particularly when relevant to a traditional cultural analysis, are also best gathered in a setting as nearly as possible approximating the ancestral one.

3.1.2 Linguistics in the field is more than linguistics The main characteristic of actual fieldwork in the field – away from the ivory tower, specifically where the linguist does not get to sleep in his or her own bed – is that it's a 24-hour-a-day operation.

When I initially agreed to write this chapter, one of the editors of this volume opined that field linguistics must necessarily involve eating weird food and developing strange illnesses. Well, of course that's true. I have eaten grasshoppers (in Oaxaca) and squirrel (in Oklahoma); I have suffered from deeply embedded ticks that had to be surgically removed (in Oklahoma) and Montezuma's revenge (in Oaxaca).

But weird food and illnesses are just part of the story, and not a very big part. What's different about fieldwork in the field is that the linguist participates in speakers' lives much more than when doing work with speakers in his or her own community.

A student of mine once drafted a small grant proposal in which she estimated her daily mileage on a field trip at twice the distance between the motel she proposed to stay at and the location at which she hoped to meet with speakers. I suggested that this did not include the inevitable mileage spent driving around trying to locate speakers, or doing other things such as taking speakers without cars to forgotten doctor's appointments. Of course this isn't the field worker's "job," but if you are there at someone's house with a car, won't you volunteer to take him or her to the clinic if there's no other way to go? Just as learning to be a good elicitor of data involves learning (or being reminded of) basic politeness, learning to be a successful field worker means being willing to participate. It means not assuming that it's possible to make out a schedule of field sessions in advance (so many things intervene – especially, particularly when working with elderly consultants, funerals). And it means being ready to learn about other aspects of your consultants' culture. Being willing to give up your time to do this not only will prove to be personally rewarding, but will show speakers that you are really serious about learning their language. (An excellent memoir about linguistics in the field is R. M. W. Dixon's 1984 description of *Searching for Aboriginal Languages in Australia*.)

3.1.3 Technology and the field worker When I started doing fieldwork, there were no personal computers, and if I wanted to record a speaker I had to bring along a reel-to-reel tape-recorder – even the small portable models were bigger than a fat telephone book.

The first dictionary I did (a preliminary version of my Mojave dictionary; Munro and Brown 1976; Munro, Brown, and Crawford 1992) was compiled using three-inch by five-inch slips (some linguists, I know, preferred four-inch by six-inch slips…) – not cards (too thick!), but slips of ordinary paper, which were arranged alphabetically in a file box (one hundred slips take up only a little more than half an inch…). Reluctantly, I have stopped introducing field methods classes to the joys of file slips, which I still feel are unparalleled for their ability to be freely manipulated and arranged in different ways. But I don't use paper slips much myself anymore, so it doesn't seem right to require students to make a slip file, as I once did.

Computers have changed fieldwork considerably, and they are now easily portable; with battery packs, they can be taken anywhere (and in fact solar chargers allow using them even where there is no electricity). With a portable computer, one can display and examine wave forms and

pitch tracks, add to a growing database, and easily search through previous recordings and related data. There are now intricate programs for the construction of dictionaries and text analysis, including SIL's FLex, which allows sharing of data between several analysts working simultaneously.

Tape recorders have also improved exponentially in the last few decades. Even inexpensive portable tape recorders often produce excellent recordings, especially with a good microphone. The availability of good digital recorders and microphones allows the recording of high quality data, suitable for all types of laboratory analysis, even under the most difficult field conditions. Camcorders and other video equipment allow any field worker to record gestures and other non-verbal cues, stimulating types of analysis hitherto never before attempted with exotic languages.

4 Analyzing the Data, and What to Do with It

4.1 Basic analysis

The most useful way to find out what you do not know is to try to describe what you do know. It is very important to keep writing – sections of a grammar (or dissertation), papers, anything – and to try to see how well the language can be described within the framework of what you already know about language and how it works.

Linguistic analysis of many sorts is covered in other chapters of this book. The most important point for a field linguist to remember is that analysis must be ongoing. The notion (which one sometimes hears) that a graduate student can go off to the field and collect data for a year, and then come back to the university and begin writing a dissertation, seems ridiculous to me. The only way to know for sure what you need to know next is to have tried your best to understand and analyze what you have already learned.

The minimum sort of ongoing analysis, which I recommend to all my students, is to type up reports of each field session (or, alternatively, to enter new data in some sort of database), preferably with notes, comments, and preliminary analysis. Looking critically at the data in this way helps to reveal gaps in paradigms and new directions to take in the next session.

4.2 Writing the language

An early goal in any sustained fieldwork should be to arrive at an understanding of the language's basic phonology. This is obviously easier with some languages than with others (though almost all languages present tricky analytical issues). But without knowing which sounds are contrastive and what sort of allophonic variation may occur in which environments, the linguist is apt to get bogged down in low-level phonetic transcription and to miss significant generalizations.

The particular phonetic transcription system adopted is not too important (I think), as long as it is used consistently. My colleagues who work on American Indian languages often use the "Americanist" symbols rather than the International Phonetic Alphabet (IPA) (for instance, writing *š* instead of *ʃ* for the sound of *sh* in *ship*), but neither of these has any particular advantage over the other, as long as one clearly sets out what each symbol means for the language under study.

Once the phonology is analyzed, it is worthwhile to adopt a clear phonemic orthography (if one is not already available). Using a phonemic orthography simplifies the presentation of data and makes it easier to present one's analysis in almost any forum, except for certain types of phonetic or phonological study. Failing to use a phonemic orthography (if you yourself understand the phonemic analysis) is insisting on obfuscation: you are depriving the more casual reader of knowledge you possess. (One of the classic descriptive grammars of all time is Edward Sapir's description of Southern Paiute (1930–1).[6] But few of the people who have praised this careful and

indeed beautifully complete fieldwork-based study have spent much time with it, for it is exceptionally difficult for the casual reader, since it mixes at least three levels of transcription – very abstract phonemic, fairly superficial phonetic, and extremely detailed phonetic – and is often exasperatingly hard to work through!)

I strongly recommend that field linguists – and others working with languages that do not already have an established orthography – develop not just a phonemic orthography, but a practical orthography, one that can be written entirely on a standard keyboard (in other words, one that uses no special nontypeable phonetic symbols or diacritics). Using such an orthography means that one can enter data in any computer application (including email and cell phones!) without the use of special fonts, but it has a more important practical value. Ordinary people – native speakers and their relatives, scholars in other disciplines, and interested laypeople – can easily learn to read and use an orthography that doesn't make use of special characters, but they are often mystified or even repulsed by an orthography that makes use of unfamiliar symbols. I have heard native speakers beg linguists to help them develop a way to write their languages without special symbols, but such pleas sometimes fall on deaf ears. This is odd, since the meanings of the symbols in a practical orthography can be explained just as clearly for the benefit of linguists (with a one-time use of IPA, perhaps) as other symbols can, so that everyone benefits.

Certainly, some languages are harder to devise orthographies for than others (particularly given the odd biases of current Euro-centered keyboards, which for example include *ã* and *õ*, but no comparable symbols for *e*, *i*, and *u*). But it is well worth it to put out the effort to develop such systems.[7]

4.3 Describing the language

Some field linguists learn a lot about languages they work on, but never publish anything. This is a criminal shame, especially since the languages in question may not be spoken forever. I believe that any linguist who engages in extensive fieldwork has a duty to publish (or otherwise make available) as much of his or her analysis of the language as possible. Preferably, such material should be disseminated in the form of clear description that is accessible as wide a range of readers as possible. This is particularly true of languages that are seriously endangered, for which it is (alas) relatively easy to foresee a time when today's linguistic description will be the only source of information on the language.

The late Mary R. Haas, who founded the Survey of California Indian Languages at the University of California, Berkeley, and trained several generations of field linguists, taught her students that the most important goal of the descriptive linguist should be to produce a grammar, a dictionary, and a collection of texts. Such material can serve as the basis for production of pedagogical materials for language revival, cultural enrichment materials, background research in many disciplines other than linguistics, and later comparative and theoretical linguistic research. My own teacher, the late Margaret Langdon, wrote, "Only after seriously confronting (if not completing) such a task can one call oneself a linguist. On the other hand, I am convinced that this task cannot be approached without some theoretical assumptions to guide the enterprise and to provide the questions to be answered" (quoted in Hinton and Munro 1998: 1).

There are, of course, bad and good descriptions.[8] A good description must be written with a solid understanding of the workings not just of the language being described, but also of language in general. For example, if someone making a dictionary has not worked out how many parts of speech the language has, with what morphological and syntactic characteristics, that dictionary will probably be incoherent.[9] It may come as a surprise to readers familiar only with European languages, but languages vary widely in just this regard (Munro 2006). Although (I believe) all languages have verbs, nouns, and probably a few recalcitrant other types of words often called "particles," many languages have no words corresponding to articles, and quite

a large number of languages have no adjectives or quantifiers as we understand those terms with regard to English.[10] (In many languages adjectival notions are expressed by a subclass of either verbs or nouns, and I know quite a few languages in which quantifiers clearly are verbs, taking all expected verbal inflection.) But to accurately list and define words in a dictionary, the linguist must understand what the significant syntactic and morphological oppositions in the language are, and endeavor to encode these as clearly and accurately as possible. It is perhaps because of the field linguist's inevitable preoccupation with the minutiae of describing everything, of letting no piece of data escape unrecorded, that basic description is often dismissed as "pretheoretical." This term is sometimes used by theoretical linguists to mean that a description contains nothing relevant to current theory – no new constraints, no new projections… The irony is that such description is very often used as input to new theoretical advances (as I discuss further below), but it could not (or should not) be so used if it were not rigorously presented.

5 Contributions of Field Linguistics to Linguistic Theory and other Scholarly Work

Basic descriptive data and analysis by field linguists contributes to the development of linguistic theory in two principal ways.

First, good description advances the theory by "testing" it, examining the way in which new data can be presented within current models, and showing how those claims must be extended and modified to handle new facts. Perhaps the most important early example of the importance of novel field data for the development of theory is Sapir's seminal paper "The Psychological Reality of Phonemes" (1949), which established the existence of naive speakers' mental concept of the phoneme (in Southern Paiute, Sarcee, and Nootka), foreshadowing the development of generative grammar. The best example I know of of a linguist who in his own work and that of his students was constantly concerned with the relationship of field data to theory is the late Kenneth Hale, practically all of whose works present new and interesting data within a highly relevant theoretical context. Among the most significant is Hale's work (based on languages of Australia and the Americas) on the notion of nonconfigurationality, which inspired extensive work on clause structure and pronominal and other arguments. Related work by Mark Baker, based in large part on fieldwork on Mohawk, resulted in important contributions to the theoretical treatment of incorporation (1988) and polysynthesis (1996),[11] contributing to the development of the Minimalist program in syntax.

Excellent contributions to linguistic theory based on solid fieldwork abound. Fieldwork-based dissertations by six of my students, for example, offered approaches to syntactic problems involving binding in Choctaw (George A. Broadwell 1990), Wh-movement in Western Apache (Brian C. Potter 1997), antisymmetry in San Lucas Quiaviní Zapotec (Felicia A. Lee 1999), stativity and distributed morphology in Pima (Eric Jackson 2005), subjecthood in Macuiltianguis Zapotec (John Foreman, 2006), and adpositional relations in Tlacolula Valley Zapotec (Brook Danielle Lillehaugen, 2006). Each of these works – like the best such descriptions – includes descriptive sections as well as theoretical argumentation. Others produced more typologically oriented work on aspect and modality in Tolkapaya Yavapai (Heather K. Hardy 1979) and grammatical relations in Cherokee (Janine Scancarelli 1987) and more general grammars on never previously described Maricopa (Lynn Gordon 1980) and Yagua (Doris L. Payne 1985).

Differences among languages provide valuable clues to how cognitive processes are related to speech, and the goal of much theoretical linguistics is to examine this relationship. But the theory can only be truly extended as it incorporates increasingly novel data-based observations. The relevance of these observations is not always immediately appreciated, so sometimes the most

important contribution of a descriptive linguist will simply be to record facts about language that do not yet fit into any theoretical paradigm – but which will be relevant for future ones. Most typically, such pieces of data are noted by linguists without a theoretical axe to grind, whose whole purpose is to provide as complete as possible a description as possible. Such people often note the existence of phenomena that are as yet irrelevant for current theory.

For instance, descriptive linguists have noted many ways in which pronominal agreement and case systems deviate from the Indo-European nominative-accusative norm. Since the 1970s, these have become an important subject for typological analysis – Anderson's (1976) and Dixon's (1979) important studies of ergativity would have been impossible without a vast body of primary "pure" descriptions. More recently ergativity has been a concern even in highly theoretical work (e.g., by Hale and Keyser 1993; Laka 1992; Legate 2008).

Many other aspects of typological research advance through the work of much earlier descriptive linguists. When I was in graduate school it was a commonplace truism that no language had a basic word order that began with the object. SVO, SOV, VSO were commonplace basic word orders, VOS had been observed in a few languages, but OSV and OVS did not occur – of course prompting the development of typological theories to account for this observation. Even as I was being taught about this, however, field linguists associated with the Summer Institute of Linguistics were recording Amazonian languages with just this word order. They did not describe these languages with an eye toward upsetting typological claims that they may not even have been aware of; they simply wanted to describe the languages they worked on thoroughly and well. But their work led to the advancement of typological studies (e.g., Derbyshire and Pullum 1981).

The study of phonology traditionally draws on a wider linguistic data base than syntax. Although Chomsky and Halle's pioneering study *The Sound Pattern of English* (1968) is now often viewed as the epitome of abstraction, this work set an important standard in terms of the number of languages that were cited in support of its claims (and, in particular, that went into the development of its feature system). Increasing numbers of phonetic studies have made possible sophisticated surveys of a very wide range of languages (e.g., Ladefoged and Maddieson 1996), which in turn provide input into theoretical studies of all aspects of phonology. Whole subfields of phonology, such as autosegmental phonology (Goldsmith 1979), arose because of the realization that there were types of phonetic data that could not be handled easily within current theory. My students whose dissertations focused on the phonology and phonetics of Choctaw (Charles H. Ulrich 1986), Nez Perce (Harold Crook 1999), and Yalálag Zapotec (Heriberto Avelino Becerra 2004) likewise provided more input to comparative and theoretical studies.

Field linguists also contribute to other scholarly activity besides theoretical work in syntax, semantics, phonetics, and phonology. At the beginning of this chapter I mentioned several other areas of linguistics that rely on work with native speakers, such as acquisition studies and sociolinguistics. Such work cannot be done easily – or perhaps cannot be done at all – on languages for which no basic description exists, so providing basic descriptions lays the foundation for later linguistic analysis of almost any kind. Another field of linguistics for which basic description of as many languages as possible is vital is historical linguistics, and the related areas of classification and dialectology. Comparative and historical work must be based on basic field data.

Researchers in many other fields draw on primary linguistic description (and greatly appreciate it if it is as theoretically neutral and devoid of jargon as possible). Anthropologists, ethnologists, and historians make use of linguistic description for research both on contemporary populations and on historical records that may include material in minority languages. Scholars studying place names, ethnobotany, and many other aspects of culture make use of primary linguistic description, particularly in the form of dictionaries.

6 The Highest Contribution

One of the most important reasons to do primary description is to preserve languages that may otherwise pass away. Languages reflect much of their speakers' culture and experience. Much of a society's knowledge and unique expressions will inevitably be lost with the loss of language. Linguistic diversity is one of the most visible and important aspects of mankind's cultural diversity, and language loss diminishes this diversity. The passing of any language reduces the range of human expressive power, and may lessen our chance of figuring out how language is realized in the mind.

Some linguists do extensive fieldwork in graduate school, write a dissertation, and then go on to careers as professional academics, publishing only on theoretical issues. Others (like me) never recover from the bite of the fieldwork bug, and must always go on to study one more language, or to learn one more word to add to the current dictionary in progress.

The best contribution this last group of field linguists can make is to produce descriptions like those I have described here, which can be used not only as the basis of linguistic and other scholarly research, but also by the communities of the native speakers who have helped us, for assistance in language revitalization and cultural awareness programs or to promote literacy. Dictionaries that can be used by ordinary people, written with clearly explained, easily understood orthographies, and grammars (especially teaching grammars) that can be used by intelligent, motivated laypeople, are among the descriptive linguist's most useful publications.

These can also, of course, be the most enduring of contributions. Check the circulation records of any large library. With virtually no exceptions, the linguistic books that are still being checked out 30 or 50 or 70 years after they were written are basic descriptions, not theoretical tomes.

NOTES

1 The data in this chapter are cited in practical orthography (see Section 4.2). For more about Chickasaw orthography, see Munro and Willmond (1994, 2008).

2 There are especially unfortunate potential parallels between a traditional police informer and a linguistic "informant": both are paid by an outsider in authority (surely a university professor is such a person) to reveal confidential information known only to the payee's intimate circle. I believe that there are many potential nonfinancial benefits to a native speaker who works with a linguist – the work is often intellectually stimulating, the native speaker usually winds up learning interesting things about his or her language, and he may receive the gratification (perhaps with associated status or recognition) of contributing to his language's preservation. Nonetheless, however, there are certainly groups who regard the teaching of their language to outsiders as a betrayal. Why should linguists use a term that invites this suggestion?

3 Most of these sets are named for their initial consonant. The "D+" endings are a bit more complex than the set of "D" endings (though I'm open to other names!). "I" means "infixed"; "S" means "short". The set of endings called "D" here are called "R" and those called "D+" are called "D" in some literature (e.g., Munro 1997). The set of endings called "L" were recently discovered and have not been reported previously. (All these terms are used in Munro et al. (2016), which also explains the practical orthography.) Others have analyzed this system with basic pronominal suffix morphemes (similar to the S or I sets) combined with prepronominal "linking" suffixes like *-ti-*, *-di-*, *ni-*, and the like (with a rule deleting the first of two adjacent vowels), although it is difficult to suggest a meaning for these morphemes. But this still leaves other separate sets! My current feelings are certainly influenced by the fact that I have to teach this material, since in undergraduate classes a paradigmatic approach seems most appropriate pedagogically. The rules for when

the prepronominal suffixes are used would need to be pretty complex.

4 Garifuna "hit" has other forms in different syntactic contexts.

5 Garifuna has a rule by which many sequences of *a* + *i* become *ei* and sequences of *a* + *u* become *ou*. Unfortunately, it's not a regular rule.

6 Based for the most part, in fact, on data from a displaced speaker from Utah, Tony Tillohash, a student at Carlisle Indian School in Pennsylvania.

7 I discuss some of the problems of devising practical orthographies, and some clever solutions to these problems by a variety of field linguists, in Munro (1996), which incidentally presents an early orthography for San Lucas Quiaviní Zapotec, developed with my collaborator Felipe Lopez, that has now been modified several times. Cahill and Rice (2014) is a collection of very useful papers about orthographies – my own paper in that volume (Munro 2014) is about breaking orthography development rules.

8 I could say a lot about bad description, but I won't. There are good grammars and bad grammars, and good dictionaries and bad dictionaries. Usually it's pretty easy to tell the difference just by inspection – inconsistencies and things that don't make sense are pretty easy to spot if you look for them. But sometimes one can't be sure one's dealing with a bad description until one actually studies the language being described. This is scary, given that some languages can no longer in fact be studied! But the only solution is for more people to try to do the best job with description that they can.

9 A colleague once said (seriously, I believe, at the time) that you don't even need to be a linguist to make a dictionary; all you have to do is write down words. This ignores the points just made in the text, as well as the need for a thorough phonological (and orthographic) analysis of the sort described earlier. I think that most likely this colleague later revised this view, and perhaps spoke hastily even on this occasion. However, such remarks illustrate the relatively low standing of descriptive linguists in our field. (Frawley, Hill, and Munro (2002) is a collection of papers on the issues involved in making fieldwork-based dictionaries.)

10 Perhaps it is true that there are indeed languages for which there really is no distinction between nouns and verbs, but I have no personal experience with such languages.

11 Polysynthetic languages are those that express many meanings within a single verb word. Baker's definition of polysynthesis is more restricted than the usual understanding of this term, consequently (from my point of view) making the term less useful and interesting. But his claims based on his notion of polysynthesis are provocative and important.

REFERENCES

Anderson, S. R. (1976). On the notion of subject in ergative languages. In C. N. Li (ed.), *Subject and Topic*. New York: Academic Press.

Avelino Becerra, H. (2004). *Topics in Yalálag Zapotec, with Particular Reference to its Phonetic Structures*. PhD dissertation, UCLA.

Baker, M. (1988). *Incorporation: A Theory of Grammatical Function Changing*. Chicago, IL: University of Chicago Press.

Baker, M. (1996). *The Polysynthesis Parameter*. New York: Oxford University Press.

Broadwell, G. A. (1990). *Extending the Binding Theory: A Muskogean Case Study*. PhD dissertation, UCLA.

Cahill, M. and Rice, K. (eds) (2014). *Developing Orthographies for Unwritten Languages*.. Dallas, TX: SIL International Publications.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. New York, Evanston, IL, and London: Harper and Row.

Crook, H. (1999). *The Phonology and Morphology of Nez Perce Stress*. PhD dissertation, UCLA.

Derbyshire, D. C. and Pullum, G. K. (1981). Object initial languages. *International Journal of American Linguistics*, 47, 192–214.

Dixon, R. M. W. (1979). Ergativity. *Language*, 55, 59–138.

Dixon, R. M. W. (1984). *Searching for Aboriginal Languages: Memoirs of a Field Worker*. Chicago, IL, and London: University of Chicago Press.

Frawley, W. J., Hill, K. C., and Munro, P. (eds) (2002). *Making Dictionaries: Preserving Indigenous Languages of the Americas*. Berkeley, CA: University of California Press.

Goldsmith, J. (1979). *Autosegmental Phonology*. New York: Garland.

Gordon, L. (1980). *Maricopa Morphology and Syntax*. PhD dissertation, UCLA.

Hale, K. and Keyser, J. (1993). On argument structure and the lexical expression of syntactic relations. In

K. Hale and S. J. Keyser (eds), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* (pp. 53–108). Cambridge, MA: MIT Press.

Hardy, H. K. (1979). *Tolkapaya Syntax: Aspect, Modality, and Adverbial Modification in a Yavapai Dialect*. PhD dissertation, UCLA.

Hinton, L. and Munro, P. (eds) (1998a). *Studies in American Indian Languages: Description and Theory*. University of California Publications in Linguistics 131. Berkeley, CA: University of California Press.

Hinton, L. and Munro, P. (1998b). Introduction. In Hinton and Munro 1998a, pp. 1–4.

Jackson, Eric (2005). *Resultatives, Derived Statives, and Lexical Semantic Structure*. PhD dissertation, UCLA.

Ladefoged, P. and Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Laka, I. (1992). Ergative for unergatives. UCLA Colloquium.

Lee, F. A. (1999). *Antisymmetry and the Syntax of San Lucas Quiaviní Zapotec*. PhD dissertation, UCLA.

Legate, J. A. (2008). Morphological and abstract case. *Linguistic Inquiry*, 39, 1, 55–101.

Munro, P. (1988). Diminutive syntax. In W. Miller (ed.), *A Festschrift for Mary R. Haas* (pp. 539–56). The Hague: Mouton.

Munro, P. (1996). Making a Zapotec dictionary. *Dictionaries*, 17, 131–55.

Munro, P. (1997). The Garifuna gender system. In J. H. Hill, P. J. Mistry, and L. Campbell (eds), *The Life of Language: Papers in Honor of William Bright*. The Hague: Mouton de Gruyter.

Munro, P. (1998). Chickasaw expressive "say" constructions. In Hinton and Munro 1998a, pp. 180–6.

Munro, P. (2006). From parts of speech to the grammar. *Perspectives on Grammar Writing*. Special issue of *Studies in Language*, 30, 2, 307–50. (Reprinted in T. E. Payne and D. J. Weber (eds) (2007), *Perspectives on Grammar Writing* (pp. 71–112). Amsterdam: John Benjamins.

Munro, P. (2014). Breaking rules for orthography development. In Cahill and Rice 2014, pp. 169–87.

Munro, P. and Brown, N. (1976). *A Mojave Dictionary*. Privately printed and distributed, partly under the auspices of the CRIT Museum, Colorado River Indian Reservation, Parker, AZ.

Munro, P., Brown, N., and Crawford, J. G. (1992). *A Mojave Dictionary*. UCLA Occasional Papers in Linguistics 10. Los Angeles, CA: Linguistics Department, University of California.

Munro, P. and Lopez, M. with Lambey-Martinez, A., and Martinez-Ciego, V. M. (2016). *Adímureha waman Garífuna (Let's Speak Garifuna)*. Westwood, CA: UCLA Course Reader Solutions.

Munro, P. and Willmond, C. (1994). *Chickasaw: An Analytical Dictionary*. Norman, OK: University of Oklahoma Press.

Munro, P. and Willmond, C. 2008. *Let's Speak Chickasaw: Chikashshanompa' Kilanompoli'*. Norman, OK: University of Oklahoma Press.

Payne, D. L. (1985). *Aspects of the Grammar of Yagua: A Typological Perspective*. PhD dissertation, UCLA.

Potter, B. (1997). *Wh/Indefiniteness and the Structure of the Clause in Western Apache*. PhD dissertation, UCLA.

Sapir, E. (1930–1). The Southern Paiute language. *Proceedings of the American Academy of Arts and Sciences*, 65, 1–3.

Sapir, E. (1949). The psychological reality of phonemes. In *Selected Writings of Edward Sapir in Language, Culture, and Personality*, ed. D. G. Mandelbaum (pp. 46–60). (Originally published in French in 1933 as La Réalité psychologique des phonèmes, *Journal de Psychologie Normale et Pathologique*, 30, 247–65.)

Scancarelli, J. (1987). *Grammatical Relations and Verb Agreement in Cherokee*. PhD dissertation, UCLA.

Ulrich, C. H. (1986). *Choctaw Morphophonology*. PhD dissertation, UCLA.

5 Writing Systems*

PETER T. DANIELS

When the first edition of this chapter was prepared, some fifteen years ago, it was necessary to say that chapters on writing systems were very rare in surveys of linguistics, and only two could be listed. The "surveys of linguistics" industry seems to have petered out since then, in favor of major series of "Handbooks" and "Companions" on specific fields, such as that which includes the present volume. But a salutary simultaneous development has been the inclusion of substantive chapters on writing systems in basic textbooks (Daniels 2010).

For a century or so – since the realization that unwritten languages are as legitimate a field of study, and perhaps a more important one, than the world's handful of literary languages – writing systems were (rightly) seen as secondary to phonological systems and (wrongly) set aside as unworthy of study or at best irrelevant to spoken language. The one exception was I. J. Gelb's attempt (1952) to create a theory of writing informed by the linguistics of his time. Gelb said that what he wrote was meant to be the first word, not the last word, on the subject, but no successors appeared until after his death in 1985.[1] Although there have been few linguistic explorations of writing, a number of encyclopedic compilations have appeared, concerned largely with the historical development and diffusion of writing,[2] though various popularizations, both new and old, tend to be less than accurate (Daniels 2000). Daniels and Bright 1996 (*The World's Writing Systems (WWS)*) includes theoretical and historical materials but is primarily descriptive, providing information (not previously gathered together) on how most contemporary and some earlier scripts represent (the sounds of) the languages they record.

Space limitations preclude reprising the catalogue of writing systems included in the first edition (for which see note 2, and Daniels 2009, albeit without illustrations). In its stead, a few examples illustrate the principles of script inventions and adaptations. Further examples and detail will be found in Daniels (in press).

* I am grateful for the comments and suggestions of the volume editors, and of Jerrold S. Cooper, Victor Mair, David L. Share, P. Oktor Skjærvø, Ben Zimmer, and the late John DeFrancis and M. O'Connor. Space limitations preclude incorporating them all, which can only be to the detriment of the chapter.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

1 The Diversity of Writing Systems

The key to the nature of writing is the primacy of the syllable. Psycholinguists find that people not literate in an alphabetic script are unable to manipulate portions of the speech stream at the level of the segment (Daniels 1992b); educational psychologists find that syllabic approaches to teaching children to read can be more successful than approaches requiring them to identify subsyllabic segments (overview in Share and Daniels 2015); phonologists increasingly work with levels of analysis other than that of the segment or individual sound (but none seems to have broken entirely with the Cs and Vs of alphabet-based analysis). The inventions by untutored writing-inventors record syllables.

The syllabic orientation to the study of writing systems leads to a five-way typology, which is especially useful in investigating the historical development of writing. In the order they came into being, the five types of writing system are: *logosyllabary* (more precisely *morphosyllabary*), in which each character stands for a morpheme, and the characters can be used for the sound of the morpheme as well as for its meaning (in C. F. Hockett's formulation: "unit symbols represent syllables but with homophones distinguished" 1997: 381 – there can be no purely logographic script); *syllabary*, in which each character stands for a syllable; *abjad* (the Semitic-type script), in which each character stands for a consonant; *alphabet* (the Greek-type script), in which each character stands for a consonant or a vowel; and *abugida* (the Sanskrit-type script), in which each character stands for a consonant accompanied by a particular vowel, usually /a/, and the other vowels (or no vowel) are indicated by consistent additions to the consonant symbols. For an introductory purpose, however, it is more convenient to take up the various types in a different order.

1.1 Syllabaries

The earliest known and most celebrated invention of a writing system in modern times by someone who knew nothing of writing beyond observing others using it to communicate at a distance – well over a dozen have been described (Schmitt 1980) – is that for Cherokee, devised by Sequoyah (ca. 1785–1843) around 1821 (Foreman 1938). After some unsuccessful attempts to come up with a symbol for every word in the language, he realized that every word contained identifiable parts. The parts he could identify were *syllables*: what others, those accustomed to writing with an alphabet, would consider sequences of a consonant and a vowel. The script includes 85 such characters. Some examples: ᏣᎳᎩ *tsalaki* 'Cherokee'; ᎠᏕᎶ *ama* 'water' or *āma* 'salt'; ᏫᏚᏢ *khāka'li* 'February'; but ᎪᎳ *kōla* 'winter' or *khōla* 'bone'.

These examples demonstrate a feature often found in writing systems: the omission, either systematic or practical, of certain phonemic information. Vowel length is significant in Cherokee, but no provision is made for expressing it or the glottal stop; and all obstruents can be aspirated or not, but the distinction is written only in a few frequent syllables: *kha* and *ka* are written differently, but *ko* and *kho* are not (Scancarelli 1992). Nor is any provision made for notating tone.

Also visible here is a feature unique to Cherokee among freshly invented scripts: the resemblance of some characters in shape, though never in sound, to letters of the roman alphabet. Alongside the Ꭺ, Ꮆ, Ꭰ, Ꮃ, and Ꭹ already seen, there are characters like Ꭱ *e*, Ꭲ *i*, Ꭵ *to*, Ꮤ *tha*, and Ꮓ *no*. Sequoyah's original designs for the characters were more elaborate, involving many graceful flourishes of the pen. It was long thought that the present shapes were introduced by the first person to print Cherokee, in 1828, to take advantage of the type already available to printers, but some handwritten examples of the simpler shapes have now been found that predate by several years any thought of printing Cherokee (Walker and Sarbaugh 1993).

The other modern syllabary that is in widespread use among the speakers of its language is that of the Vai people, whose territory straddled the border of what are now Liberia and Sierra Leone, on the west coast of Africa (Scribner and Cole 1981). Its inventor, Mumɔlɔ Duwale Bukɛlɛ,

probably had a few years of elementary education in a British missionary school, but when in 1833 or 1834 he resolved to create a script for his people different from that of either the Christian or the Muslim scriptures that were revered in the region, he chose to make it a syllabary. There is a slim possibility that he had heard of Sequoyah's achievement from American missionaries (Tuchscherer and Hair 2002).

Over the subsequent decades, a popular employment for Vai men seems to have been as mercenaries, and it is possible that it was their example that led a dozen or more visionaries in West Africa to create syllabaries for their own languages and peoples, though none of those later systems exhibit visual similarity to Vai (a number of them are attractively presented in Mafundikwa 2004).

While some or all of the modern-day West African syllabaries may have come about by such "stimulus diffusion,"[3] this is unlikely in the extreme for syllabaries like the Caroline Islands (Micronesia, North Pacific) and the Alaska. What all these examples show is that there is something fundamental, something primal, about the syllable, especially the CV syllable. While articulatory phoneticians cannot agree on what makes a syllable, acoustic phonetics reveals that there are no segment Cs or Vs in the stream of speech: syllable-initial consonants are, essentially, deformations of the "formants" – resonances created by the cavities (pharynx, mouth) in the vocal tract – recognition of which is what produces the identification of the vowels of a language (see Chapter 9).

1.2 Alphabets

The English alphabet is familiar to everyone reading this book, but it is not the most convenient place to begin a discussion of alphabets – because of the *ghoti* phenomenon: the claim (which despite popular rumor has no connection whatsoever with George Bernard Shaw [Zimmer 2010]) that the word *fish* can be spelled *gh-* as in *enough*, *-o-* as in *women*, *-ti* as in *nation*. Instead, we can begin with the first alphabet and return to *ghoti* later.

The first alphabet was that for Greek. There were numerous local varieties (Jeffery 1990), not surprisingly for an innovation just coming into use; but that of Athens, which was officially adopted in preference to all others in 402/401 BCE, comprised the 24 letters still used today. Even then, that choice was not ideal – long and short /e/ and /o/ have separate letters, but long and short /a/ do not; nor is there a letter for /h/, which was found in many Greek dialects other than that of Athens. What makes the Greek script an alphabet is its isolation of vowels and consonants, and its lack of distinction in the treatment of them. There are letters for A, E, H (*ē*), I, O, Y (*u*), and Ω (*ō*) – and the very presence in this list of H and Y shows (by comparison with the English alphabet) that you can't tell by looking at them which are vowels and which are consonants.

Usually, when an alphabet is first devised for a particular language, there is a one-to-one correspondence between the letters of the alphabet and the phonemes of the language.[4] Unfortunately for this "puristic" view, which for "spelling reformers" is the only legitimate relation between sound and spelling, languages are constantly changing whereas once something is written down, it's permanent, and if readers are to be able to read documents written even a few decades earlier, they need to be able to correlate the sounds of their language with the way the language was spelled, perhaps representing somewhat different sounds, at an earlier period. Over time, there have been two or three ways of dealing with this "problem." One is to ignore it, and to continue whatever spelling habits were in force when a language was first written down.[5] The other is to periodically adjust the spelling conventions to more closely reflect the phonological reality of the language.[6] In the case of Greek (and of German), such decisions are made by governmental authorities. Greek B at first represented /b/ (as its descendant does in Latin) but already in the first millennium CE, thanks to sound change, it represented /v/ (as its descendant does in Russian). Modern Greek, though, once again has /b/, and the sequence of letters MB is used for /b/.

But for English, there is not and never has been any sort of governmental authority to impose such a change. Instead, orthographic standards are set by influential individuals: in the sixteenth and seventeenth centuries, these were mostly Classical scholars attempting to regularize and improve English spelling, often with conflicting recommendations (Brengelman 1980); but in 1755, when Samuel Johnson completed his *Dictionary*, with its not always successful striving for full consistency, it almost instantly became the standard for English orthography. Noah Webster attempted to differentiate American spelling in his two dictionaries – a small one in 1806, a comprehensive one in 1828 – and succeeded with some of his choices (*color* for *colour*, *music* for *musick*, *meter* for *metre*), but not with many others (e.g., *opake* for *opaque*).

Webster didn't even discuss the pronunciation of *gh* or *ti* alongside his other recommendations (in the Preface to Webster 1806, vi–x), though s.v. "women" (352) he provides the note "*The primitive and correct orthography is wimmen.*" The likeliest explanation for the truly anomalous spelling *women* is that *woman* adopted the relic ablaut pluralization seen in *foot/feet* or *goose/geese*, with a high front vowel, but did so after the spelling had become fixed. In turn, the *o* displays a purely graphic phenomenon seen in words like *womb* and *monk*: the expected spelling *u* for the vowel would result in a picket-fence of vertical strokes (wumb, munk) that could be somewhat relieved by an o instead (womb, monk).

The digraph *gh* illustrates two less unusual sources for apparent "lack of fit" between an alphabet and a language's phonology: sound change, and foreign influence. Old English had a velar fricative /x/. By Middle English times, the few remaining instances of the voiced variety [ɣ] were sometimes spelled using the old Irish form of *g*, *ȝ* (known as "yogh"). Both the letter and the sound were unfamiliar to the Norman scribes who were accustomed to writing Latin and French. The letters for the similar sounds *g* and *h* already had other values, so they replaced *ȝ* with *gh* (matching their use of *th* for older *þ* and *ð*, and *sh* for older *sc* – though *ſ* hung on until the early nineteenth century for *s* when not ending a syllable). In various dialects and various phonological environments, [ɣ] changed in various ways, and in *enough* and a few other words, it ended up as /f/ – but only after a vowel. And then, along came the first English printer, William Caxton, who had learned his trade in Ghent and imported not only presses and type in 1475, but also Flemish workers familiar with the equipment. They tended to bring spelling habits familiar from their own language with them, and one of those was that [g] was spelled *gh* in some circumstances, and their practice has survived in a handful of English words, including *ghost*.

Finally, *ti* represents a synchronic assimilatory sound change, and a morphophonemic spelling. In such pairs as *natal* and *nation*, *construct* and *construction*, we see that the /y/ at the beginning of the suffix *ion* (cf. *domain/dominion*) palatalizes the preceding alveolar (cf. *process/procession*) when the stress falls on the preceding syllable. Retaining the *t* preserves the visual integrity of the morpheme, an orthographic property that is highly favored for inflected languages suffused with morphophonemic processes.

Thus the only possible English pronunciation for *ghoti* is as a homophone for *goaty*.

What else is special about the Greek alphabet? It influenced Western Civilization in two ways. First, Greek entrepreneurs founded colonies throughout the Mediterranean just about when they learned to write (which probably happened early in the eighth century BCE), and the various Greek dialects used some letters in different ways and tacked on extra letters as needed. From one of those colonies off the west coast of central Italy, the Etruscans adopted an alphabet that didn't have any extra letters, and the Latins (later known as Romans) adopted it from them. In Roman hands, the letters assumed the shapes of today's capital letters, and the Roman Church (insisting on Latin for its own use everywhere) carried it throughout Western Europe, where scribes used it for their own languages as well. The minuscule ("lowercase") shapes developed gradually as the need for speed and economy in writing materials led to reduction and compression.[7]

Second, in contrast to the Roman Church, the Greek-speaking Eastern churches preferred to translate their writings into local languages, and local peoples preferred to have distinctive alphabets of their own. The first to emerge was the Coptic, for the last stage of the ancient Egyptian

language. The fourth-century Gothic bishop Wulfila adapted the Greek alphabet for his East Germanic language. The Armenian saint Mesrop, in the early fifth century, is credited with no fewer than three alphabets for peoples of the Caucasus: Armenian, Georgian, and Udi (before the recent identification of its language, the scanty remains were known as "Caucasian Albanian"). And in the ninth century, the missionary brothers Constantine (later Cyril) and Methodius created the Glagolitic alphabet from Greek minuscules for an early Slavic language, and in their wake the Cyrillic alphabet was based on Greek majuscules. This is the alphabet of Russian and many languages, of various families as well as Slavic, that came under the influence of the Russian and Soviet empires.

Finally, the names of the first two Greek letters, *alpha* and *beta*, provide the word *alphabet*.

1.3 Abjads

Similarly, *abjad* is an Arabic word containing the first four letters of the Arabic script (when they're recited in the ancestral order that is still used when the letters serve as numerals), and I have adopted it to name the type of writing system whose letters denote only consonants. Again, it is more convenient to begin with the earliest rather than the most widespread example (Naveh 1987).

Phoenician was the language of a seafaring people, based at the coastal towns of Tyre and Sidon in present-day Lebanon, who flourished throughout the 1st millennium BCE and whose Punic descendants troubled Rome into the Common Era. Their script was ancestral to and comprised the same set of letters as the Hebrew script, and Phoenician examples are traditionally transliterated into Hebrew for convenience. This warning is found incised into the shaft of the earliest known Phoenician burial chamber, that of King Aḥiram of Byblos from about 1000 BCE (compare the biblical name Hiram of Tyre): לדעת : הן יפד לך : תחת זן (*read right to left*). This transliterates[8] as *ldʕt . hn ypd lk . tḥt zn* (*read left to right*), and any potential tomb-robber would have no trouble understanding that as *lidaʕt hin yupād lika taḥt zin* 'Attention! Behold, grief-shall-come to-you below here!' That interpretation involves considerable guesswork for us,[9] but as the excavator who discovered it in 1924 noted, the graffito uses the same letter shapes as the inscription on the sarcophagus beneath, indicating that literacy was available to more people than just professional scribes. To read writing is a comfortable activity for people who speak the language involved and have sufficient practice in it: other realizations of the syllables may have been possible in other contexts, but words don't generally occur in isolation with no context at all.

Preliterate peoples in ancient times seem to have thought that writing was a useful technology, and it quickly spread beyond most places where it was introduced. But communication between peoples was not all that common, and the shapes of letters used in individual communities tended to diverge from the model and from each other, and over time within communities. Such differences are useful to today's scholars in determining the place and era of inscriptions and manuscripts, but they make no difference to the reader; but one innovation in a particular corner of the abjad world did: the Aramaic languages, closely related to Phoenician, arose in the present-day Syria area. The letters of their version of the abjad were more rounded ("cursive"), but the Aramaic contribution to the development of orthography resulted from the aforementioned tendency toward conservatism in spelling. The diphthongs [aj] and [aw] changed to [eː] and [oː], but the letters י and ו for the now-missing /y/ and /w/ were still written where they had occurred. Thus some י's and ו's came to represent /ē/ and /ō/, but most י's and ו's still represented /y/ and /w/. Eventually this practice extended to the letters ה *h* and א *ʔ* for /ā/ in various places.

The fullest development of this system is seen in the Aramaic literary language Syriac, which is used to the present day in a number of Eastern churches. Here is Genesis 1:1: ܒܪܫܝܬ ܒܪܐ ܐܠܗܐ ܐܬ ܫܡܝܐ ܘܐܬ ܐܪܥܐ *bršyt brʔ ʔlhʔ yt šmyʔ wyt ʔrʕʔ = bᵊrāšīṯ bᵊrā ʔalāhā yāṯ šᵊmayā wᵊyāṯ ʔarʕā* 'in-beginning created God DIRECT-OBJECT heaven and-D.O. earth'. In this passage, the first two ܝ (*y*)'s mark the vowels, and the others, like the ܘ (*w*), are consonantal.

Here's the same verse in the Hebrew original: בראשית ברא אלהים את השמים ואת הארץ *brʔšyt brʔ ʔlhym ʔt hšmym wʔt hʔrṣ = bərêšîṯ bārâ ʔĕlohîm ʔēṯ haššāmáyim wəʔēṯ hāʔáreṣ*.[10] Hebrew, too, uses some letters to identify vowels – and did so well before adopting an Aramaic script; they are marked with circumflex accents rather than macrons in the transcription. It is often claimed that the nature of Semitic morphology, with "consonantal roots" carrying all the lexical meaning of words and "vocalic patterns" signaling only grammatical relations, was especially congenial to a consonants-only orthography; but on the one hand the "root/pattern" analysis is lamentably simplistic; and on the other, Arabic script (also a descendant of the Aramaic abjad) served both Ottoman Turkish and Classical Persian (neither of them a Semitic language) successfully for centuries.

Over time, Syriac borrowed Iranian vocabulary from the Persian administration of its territory and Greek vocabulary from Christian philosophers. The vocalization of these Indo-European words was less predictable from the syntax than is the case for Semitic words, leading to fuller use of vowel letters, sometimes even for short vowels. But again over time, the spoken languages diverged from the written languages of the sacred scriptures of all three "peoples of the Book" – Hebrew, Syriac, and Arabic. It was deemed essential to incorporate a record of the exact vocalizations into the texts, but it was equally essential not to alter the written texts as received, such as by inserting full-fledged letters for the vowels on the Greek model. First Syriac, then Arabic, and then Hebrew received supplementary systems for this purpose (Morag 1961). In Syriac, the vowel "points" gradually emerged, beginning in the sixth century CE, from single dots that marked syntactic divisions and coarser degrees of vowel quality: ܒܪܫܝܬ ܒܪܐ ܐܠܗܝܡ ܐܬ ܫܡܝܡ ܘܐܬ ܐܪܥܐ. For Hebrew, several groups of scholars devised different ways of marking vowels, with those of Tiberias eventually prevailing: בְּרֵאשִׁית בָּרָא אֱלֹהִים אֵת הַשָּׁמַיִם וְאֵת הָאָרֶץ. For Arabic, the circumstances were a bit different: the established consonantal text represented a dialect that had lost certain phonological and morphological features, and they are restored in the vocalization. The first verse of the Qur'ān is اَلْحَمْدُ لِلّٰهِ رَبِّ الْعٰلَمِيْنَ اَلرَّحْمٰنِ الرَّحِيْمِ *alḥamdu lillāhi rabbi lʕālamīn arraḥmāni rraḥīmi* 'the-praise to-Allah Lord-of the-worlds the-gracious the-merciful'. In the western part of the Syriac realm, some late manuscripts were vocalized with tiny Greek vowels above or below the line of consonants, and the shapes of the letters had developed somewhat differently: ܒ̇ܪ̈ܫ̈ܝ̇ܬ ܒ̈ܪ̈ܐ ܐ̈ܠ̈ܗ̈ܝ̇ܡ ܐ̈ܬ̈ ܫ̈ܡ̈ܝ̇ܡ ܘ̈ܐ̈ܬ̈ ܐ̈ܪ̈ܥ̈ܐ.[11] This was the first form of Syriac script to come to the attention of Europeans, so somewhat misleadingly it became widely used in students' textbooks.

In all these cases, however, the vowel pointing is normally used only for the sacred text itself; all other writings in the traditional languages omit it, except for poetry, proper names, and children's instructional material. Many modern languages that have adapted Arabic script, though, especially in Africa (Mumin and Versteegh 2014), never omit vowel notation, and this is also the case for the "Modern Syriac" koine used in the present by a wide variety of Modern Aramaic-speakers.

1.4 Abugidas

The principal home of the abugida is South Asia, although the name is from the Classical Ge'ez and Amharic languages of Ethiopia.[12] In an abugida,[13] each letter represents a consonant followed by the vowel /a/ (various modern languages of India realize it as [ə], [ʌ], or [ɔ]), or an individual vowel (these vowel letters occur only word-initially). The other vowels are written with additions to the C*a* letters. This sort of writing system spread throughout South and Southeast Asia, diversifying greatly in shape, with Buddhism and Hinduism and with Indian merchants. The example most familiar to the Western reader is probably the Devanagari script, which is used for Hindi, Marathi, and other Indic languages, and is also most common for Sanskrit as studied in the West (Sanskrit can be and is written with any of the local scripts of India). All such scripts descend from Brahmi, a script used by Emperor Aśoka (ca. 250 BCE), almost everywhere in his realm, for inscriptions publicizing his edicts in regional Indic languages known as Prakrits.

In Devanagari, क is *ka*, के is *ke*, कि is *ki*, को is *ko*, and कु is *ku*; some other choices are का for *kā*, की for *kī*, and कू for *kū*. Significant for the nature of writing systems and for the skill of reading is that the vowel additions can occur above, left of, right of, or below the consonant, while the vowel is always pronounced after the consonant (Nag 2014). A Hindi example is भारत *Bhārat* 'India'; note the ा for *-ā* after the opening भ *bh(a)*. Gujarati is very like Devanagari without the top bar: ભારત્. In three of the other beautiful Indian scripts, the same word is ভারত (Bengali), ଭାରତ (Oriya), or ഭാരത (Malayalam).

Sanskrit has closed as well as open syllables. Using a syllabary, there are two options for notating them: either the final consonant can simply be omitted, or the final consonant can be written with a letter for the consonant plus an arbitrary vowel that is skipped over when reading. But in the abugidas of India, two devices are used. At the end of a word, a mark is added; *Bhārat* in Sanskrit should be written भारत्. Within a word, however, successive consonants are combined into a single symbol. The third word of the familiar mantra *ōm mani padme hūm* is written पद्मे, where द्म is द *d(a)* reduced and followed by म *ma* (with े *e* on top). Different scripts combine letters in different ways; the equivalents in the ones seen above are પદ્મે (પ + દ્ + મ + ે), পদ্মে (প + দ্ + ম + ে), ପଦ୍ମେ (ପ + ଦ୍ + ମ + େ), and പദ്മേ (പ + ദ് + മ + േ). Bengali, though calligraphically rather different, is quite similar to Devanagari. Oriya usually subordinates the second consonant to the first. Malayalam, having undergone a twentieth-century reform in response to the introduction of the typewriter, by and large strings its letters out horizontally, using ് to elide the intervening vowel.

In the Ethiopic script, on the other hand, there is no special way to indicate that a consonant is not followed by a vowel. The seven "orders" are ከ *kä* ኩ *ku* ኪ *ki* ካ *ka* ኬ *ke* ክ *kɨ* ኮ *ko*, and the sixth order indicates either the high central vowel or vowellessness,[14] as in Ge'ez ወሀሎኩ : ውስተ : ባቢሎን : *wähällóku wìstä Babílon* 'and-I.was in Babylon'. This example also shows that neither accent nor consonant length, important to the morphology, is notated.

The clear similarity in principle between Ethiopic and Indic writing systems strongly suggests that the latter (known from at least the third century BCE) influenced the former. Ethiopic consonant letters are simply the South Arabian variety of the abjad, which evolved independently of its Phoenician cousin. The vocalization appears suddenly, and fully worked out, in inscriptions of King Ezana (mid fourth-century CE) that are simultaneous with his conversion to Christianity. The notion of adding vowel notation to a script, especially in the form of appendages, cannot have come from southwest Asian missionaries – even Syriac had nothing of the sort yet – or from Coptic ones from Egypt, who would have brought the Greek idea of vowel letters; which leaves Christian Indians (missionaries? semiliterate seamen?) crossing the Arabian Sea with traders who followed the recently discovered monsoon winds (Daniels 1992a).

Another potential influence has just been recognized. We now know that the Meroitic script, used in northern Sudan from about the third century BCE to the fourth century CE, was an abugida,[15] and during the second half of its existence, the first consonant in a cluster was written with the letter identified as /e/ (Rilly and de Voogt 2012). Whether contact between the cultures at the relevant time is plausible is an open question.

1.5 Morphosyllabaries

Only one morphosyllabic script remains in use today, the Chinese script, in two very different writing systems, the Chinese and Japanese. Almost every Chinese character – as the individual units of the script are called – consists of two components, one (called the *radical* or the *semantic*) giving a clue to the meaning of the morpheme it represents, the other (called the *phonetic*) giving an indication of the pronunciation (Qiu 2000). This indication is more or less precise in this or that character, but when the system was last codified, some two thousand years ago, the pronunciations were quite accurate. As has happened so often, the language changed, but the script and orthography did not.

At some time earlier than the earliest surviving Chinese texts, which are from ca. 1250 BCE, the ancestors of the character components were *pictograms* – stylized representations of objects named by the morphemes notated. Probably over a short period of time, the pictures devolved into patterns of brushstrokes, the number and order of writing of which are rigidly fixed; those earliest examples are incised on animal bones used in communicating with the gods (hence "Oracle Bone Inscriptions") and few of even those early forms are recognizable pictures. The characters were standardized to fit into uniform squares, no matter how many strokes they involved. Over the centuries, tens of thousands of different characters have been devised, but 5,000 or so are more than enough for most purposes. Their construction from just a few hundred components means this is not nearly so overwhelming as the figure of 5,000 might suggest.

Here are some examples of characters and their components. The radical 人 'man' appears (in its condensed form 亻) in morphemes like 仃 *dīng* 'left alone' and 仜 *hōng* 'paunch'; the radical 手 'hand' appears (in its condensed form 扌) in morphemes like 打 *dǎ* 'to strike' and 扛 *káng* 'to bear'; conversely, the phonetic 丁 *dīng* appears in 仃 and 打, and the phonetic 工 *gōng* appears in 仜 and 扛. The sounds of the morphemes represented by the radicals, and the meanings of the subset of phonetics that independently represent morphemes, are irrelevant to the characters they appear in.

Chinese dictionaries are arranged according to the radicals – 214 are recognized in the traditional count, fewer in some modern works – with the radicals in the order of the number of strokes they contain (the order of radicals within any one count is arbitrary, and different orders are found). Under each radical, the characters are organized by the total number of strokes including the phonetic. And since the phonetic is often itself an independent character, the stroke counts can grow quite large. The champion in actual use is 32-stroke 鬱 *yù* 'to implore'.

An opportunity for script reform arose with the success of the People's Revolution. Lest all connection with past culture be lost, the orthography did not change, but about 2,000 characters received "Simplified" forms: many of them were extant cursive or handwritten forms that now became standardized, but in the process, account was not taken of the component structure, so that formerly related characters lost their similarity, and the clues to meaning and pronunciation no longer existed: 變 *biàn* (with the radical 言 'speech' at the top center) 'to change' becomes 変 (with no inherited radical). The "Traditional" characters remain in use in Hong Kong and Taiwan. Also in the 1950s, in the People's Republic an auxiliary romanization system, *pinyin*, was adopted, in preference to the Western scholarly systems that had prevailed, so that "Mao Tse-tung" became "Mao Zedong," and bilingual dictionaries are nowadays ordered by romanization rather than in stroke-count character order.

Chinese civilization was immensely influential over surrounding areas, with its writing system imitated or adapted by a number of neighboring peoples. The only such surviving writing system is the Japanese (Frellesvig 2010), where characters (about 2,000 are standard) have multiple interpretations – both as the words borrowed along with them from Chinese, and also as native Japanese words or parts of them. In the latter cases, there is no relation between form and sound, so they are true morphograms. Quite early on, Japanese scribes began using a small group of characters for their sound only, and over the centuries these were standardized, their shapes simplified, into sets of 50 that represent (C)V syllables – or rather, *moras* (subunits of syllables: prolongation of a vowel or consonant, and nasalization of a vowel, also count as moras). Any Japanese text can be written just with these mora letters, called *kana* (and indeed literary works by and for women were written in this way, since women were not allowed a Classical education that would include the use of characters), but characters (called *kanji*) remain pervasive in Japanese writing. There are two sets of *kana – hiragana*, used for grammatical particles, and *katakana*, which function like English italics, for emphasis and foreign words. The words 変る KAWA-*ru* 'to change' and 変り KAWA-*ri* 'a change' contain the kanji 変 and the hiragana る *ru* and り *ri*. The borrowed word *jerī* 'jelly' is spelled in katakana: ジェリー *zi-e-ri-L*.

1.6 Informed inventions

Only two or three writing systems that, unlike those described in Section 1.1, were devised by people who were literate in an existing script are in widespread use today, and they too reveal things about the nature of writing.

The earliest is the Korean alphabet (Kim-Renaud 1997), dating to 1446 and attributed to King Sejong. Its consonant and vowel letters belong to distinctive visual sets, e.g., ㄱ *k*, ㄴ *n*, ㅅ *s*, ㅁ *m*; ㅣ *i*, ㅏ *a*, ㅡ *ŭ*, – but, in accord with the appearance of the prestigious Chinese script, the letters are combined into square syllable blocks: *ssŭgi* 쓰기 'writing'. This system did not find much use until the twentieth century, and until recently Chinese characters were included in texts, somewhat Japanese-like; but no more.

A script that has had considerable success among the speakers of three Canadian First Nations language families – Algonquian, Dene, Eskimo (Burnaby 1985) – was devised for the Algonquian language Cree by the Methodist missionary James Evans in 1841. It is known as Syllabics because, as so often, the base units represent (C)V; but, uniquely, the four vowels are denoted by the same shape in four different orientations: ᐊ *a*, ᐁ *e*, ᐃ *i*, ᐅ *o*; ᒪ *ma*, ᒣ *me*, ᒥ *mi*, ᒧ *mo*; hence ᓂᒪᒪ *nimama* 'my mother'. The final consonants of closed syllables are represented with superscripts: ᑲᐃᔑᑦᐸᐢ *ka-ishitayan* 'where I lived' (examples from the Algonquian language Ojibway).

Achieving at least some measure of acceptance is the N'ko script (Wyrod 2008), which was created in 1949 by Souleymane Kanté, a Guinean trader, Qur'ānic school teacher, and autodidact, for Maninka and is also used for several other languages of the Mande family of West Africa. It resembles Arabic script – it reads right to left, with letters within a word connected at the baseline – but all consonants and vowels are written, with tone indicators in the form of marks above the line of letters. A striking feature of N'ko orthography is that when successive syllables contain the same vowel, the vowel letter is omitted from the first of them; a true consonant cluster is indicated by an apostrophe (thus ߒ'ߎ *N'ko*: ߒߎ *nko* would be interpreted as *noko*).

A final, and rather marginal, sort of modern invention of scripts concerns those devised specifically as phonetic notations. Besides ones like the International Phonetic Alphabet, which use letters of an existing script, there are the shorthand systems formerly used by office and courtroom stenographers (both Pitman and Gregg, the leading ones of the English-writing world, are primarily abugidic, with vowels mostly omitted), and the *bopomofo* of the Republic of China, which follows Chinese phonological theory in providing letters for the *initial* consonant of each syllable, followed by the *final* (vowel and possible nasal) and the *tone*.

2 The Unity of Writing Systems

The variety of ways writing represents language might suggest that Florian Coulmas's (2005: 208) conclusion "The history of writing, therefore, cannot rely much on universal tendencies, but has to investigate the spread and transmutation of every script in its own right" could be valid. I have already mentioned, however, the fundamental consideration underlying the refutation of this statement: the primacy of the syllable. It is through the syllable that we can understand the origin of writing itself.

We must begin by defining some terms that have been used above in intuitive ways. *Writing* is *a system of more or less permanent marks used to represent an utterance in such a way that the utterance can be recovered more or less exactly without the intervention of the utterer.* Implicit in this definition is the insistence that all writing is phonologically based, as stressed by John DeFrancis (1989); excluded are what Gelb calls "forerunners of writing" (DeFrancis shows that none of the "forerunners" actually "foreran" writing) as well as all sorts of visual communication systems that are not explicitly linguistic – many writers consider any visual communication system to be "writing," but since a term is needed for those which specifically record language, why not use

the existing term in its existing meaning, and recognize that writing is one among many visible communication systems, which are *semiotic* (specifically *semasiographic*) systems?[16]

A *script* is a set of symbols (letters, characters) that represent linguistic units. The letters of the English alphabet comprise most of the letters of the roman alphabet script.[17] The Arabic script, augmented with additional consonant letters, serves many languages of the Muslim world that are unrelated to Arabic. An *orthography* is the rules relating letters to sound. English has a fairly complicated orthography; Finnish has a fairly simple orthography. A *writing system* is a combination of a script and an orthography used by one language over one period of time.

Pictograms[18] are simply stylized representations of objects, regardless of the linguistic or semiotic level at which they operate. *Ideograms* are semasiographic units that represent "ideas" rather than identifiable linguistic elements; the term is not interchangeable with "logogram."

2.1 Origin of writing

Many nonliterate peoples keep graphic records that perhaps operate on the level of the word. These records do not turn into writing, however (the "reader" cannot determine exactly what sentences the delineator had in mind). We've seen that in China, pictograms turned into logograms. This happened in two and only two other places: ancient Mesopotamia, for the Sumerian language; and Mesoamerica, either for the Mayan language family itself or for some language that preceded it in the region. It did not happen in other places where cultures had achieved similar levels of complexity, such as Indo-European– or Semitic-speaking areas, or the Quechua-speaking region of Andean South America.

The question that seems not to have previously been asked is, Why those three places, those three cultures? The question may not have been asked because writers on writing systems have usually specialized in only one of the relevant fields: West Semitic for Diringer, Cohen, and Février; cuneiform for Gelb and Friedrich; Chinese for DeFrancis; Classical for Jensen; even Mayan for Stephen Houston, who has brought together specialists from numerous fields but hasn't offered a synthesis (Houston 2004; Baines, Bennet, and Houston 2008; Houston 2012). But if we step back and look at *all* writing creations (*grammatogenies*), and if we look at them from a linguistic point of view, we see two things (Daniels 1992b).

- All new writing systems (Section 1.1) invented by nonliterates who know that writing exists are syllabaries. This suggests that syllables might be paramount in grammatogeny.
- All three languages involved in creations of writing systems from nothing share a typological similarity. They are essentially monosyllabic.

In Sumerian (Michalowski 2004), Chinese (Baxter and Sagart 2014), and ancient Mayan (Mora-Marín 2003, 2010), most morphemes are just one syllable long. Thus each pictogram representing a word also represented a single syllable, so that it was easy to reuse a pictogram to record a similar-sounding word that did not lend itself to pictography (as in the children's game of *rebus*). The Sumerian script, known as cuneiform, is unique in including VC and CVC symbols (they're called "signs" in cuneiform studies) in its inventory, as in the words *ti* 'arrow' or – by the rebus principle – 'life', *en* 'lord', and *gub* 'put'.[19] Sumerian also used *determinatives*, unpronounced signs preceding a word that indicate its semantic category – place name, wooden object, god, etc. Maya glyphs (the process of decipherment is ongoing) notate a syllable-closing consonant with a syllabogram echoing the syllable's medial vowel.[20]

The Sumerian language is agglutinative, with suffixes marking grammatical relations. In early inscriptions, such grammatical morphemes were not written.[21] Speakers of the East Semitic language Akkadian enter the historical record in the mid-third millennium BCE, and when they adopted cuneiform writing, they used the symbols primarily for their sound values (as well as for some determinatives and some logograms), and had no compunctions about writing

grammatical affixes as well as content morphemes. The Akkadian model of notating everything may have induced Sumerian scribes to record their language more fully – or the inclusion of the suffixes in the written record, like the vocalization of the Semitic abjads 2,000 years later (Section 1.3), may have reflected a need to preserve the details of a no longer spoken language.

2.2 *Diffusion of writing*

Above we saw how writing spread to neighboring peoples – around the West Semitic area (Phoenician and South Arabian to some extent, but the real winners here were the Aramaic varieties, of which nearly 50 can be identified for just the few centuries around the turn of the era [Klugkist 1982]), the Indic scripts throughout South and Southeast Asia, Chinese through East Asia, and cuneiform in Mesopotamia and then to the languages of neighboring regions, notably Hittite, Hurrian and Urartian, and Elamite. The roman alphabet and the Arabic abjad went round the world with missionaries, and the Cyrillic throughout eastern Europe and northern Asia with Russian conquest.

Writing systems "spread" through time, as well; there is chronological diversity within the unity of, for instance, English orthography, which over the ages was nourished by its Latin and Anglo-Saxon origins, Irish missionaries, French-speaking Norman conquerors, and the various "authorities" mentioned above (Section 1.2). Especially interesting are cases where cultures make adjustments to existing writing systems (as opposed to devising new scripts, as with Armenian, Georgian, or Cyrillic) to fit their own languages. Two kinds are seen in the diffusion of writing across Asia (Daniels 2007a).

2.2.1 Diffusion by adoption First, a form of Aramaic script very similar to the East Syriac (Section 1.3) and associated with the scriptures of Manichaeism came into use in eastern parts of the Iranian empires that followed Alexander the Great – successively Seleucid, Parthian, and Sassanian (which take us down to the Arab Conquest of 651). The latest variety of the script served Sogdian, an Iranian language. Sogdian script was adapted for Uyghur, a Turkic language of Inner Asia, in the eighth–ninth century (Sims-Williams 1981). Then, Uyghur script was adapted for the Mongolian language at the beginning of the Mongol Empire, in the early thirteenth century.[22] It underwent some refinements over the centuries (Kara 2005), and finally in 1599 it was adapted for the Manchu language of the Manchurian Empire. At each step, adjustments were made to bring the script in line with each language's consonant inventory – and with the Uyghur adaptation, it turned from right-to-left to top-to-bottom, likely in emulation of the Chinese style – but the orthographic principles, and hence the writing system as a whole, remained essentially the same: the Aramaic abjad traveled all across northern Asia to the Pacific coast.

2.2.2 Diffusion by adaptation The second kind of adjustment proceeded across southern Asia, but in a different mode. By 500 BCE, the Persian Empire had brought the Aramaic abjad with it. Emperor Aśoka (Section 1.4) erected bilingual inscriptions, in Aramaic and a local Prakrit, in the northwest borderlands. But the Prakrit is not simply written in a variant of Aramaic script; vowel notation was added in abugida style, and the script is called Kharoṣṭhi. What lay behind this innovation? The rich grammatical tradition associated with the name of Pāṇini was already well developed by the time writing appeared in India, and it fully understood syllables, vowels, and consonants.

Tibet, too, supported a grammatical tradition, which grew up simultaneously with the Tibetan writing system (van Schaik 2011). The Tibetan language is typologically quite different from the inflecting Semitic and Indo-European languages: like its distant relative Chinese, it is isolating, so it could be advantageous for a Tibetan script not to merge adjacent morpheme-final and morpheme-initial consonants as was done in the Indic scripts. Syllable boundaries are marked with a raised dot, while the abugidic principle of inherent basic vowel plus appendages

for other vowels continued in use. The standard example of the complexity that can be achieved in Tibetan orthography is བསྒྲུབས་ *bsgrubs* 'established', in which the principal letter is ག *ga*, with before it བ *b*, above it ས *s*, below it ྲ (the combining form of ར *r*) and ུ *u*, and after it another *b* and *s*.

Mongolian culture, on the other hand, did not support a grammatical tradition, and the morphological type differs yet again: perhaps explicit syllable boundaries are less important for an agglutinative language like Mongolian. The hPags-pa script was commissioned by Kubla Khan in 1269 to record all the languages of the Empire (it was used for exactly 100 years, and mostly for Chinese). Its lettershapes are adapted from Tibetan; with /a/ included in each consonant letter it is an abugida.[23] Its importance for the history and typology of writing, though, is that it was probably part of the input to the Korean alphabet, in which Chinese phonological theory and hPags-pa shapes were combined (Ledyard 1997).

Every script reflects some degree of "native speaker analysis" (O'Connor 1983; Daniels 2013); the lesson of the Asian sequences of transmission, as much as of the "informed" modern inventions (Section 1.6), is that real innovation in script transfer must be primed by grammatical understanding of the language that is to be written – metalinguistic knowledge of one's language: the result of deep study, not simple copying.

2.2.3 Diffusion by misunderstanding The three modes of script transmission sketched above – general neighborliness,[24] official adoption, and grammarians' insights – postdate and don't account for the three transitions that actually led to the vast majority of writing systems, all those that emerged from the Aramaic and Phoenician abjads. We must recognize three episodes of *misunderstanding* that led to revolutionary innovations in writing.

First, if the "monosyllabic" theory is correct, it is apparent that Egyptian hieroglyphs cannot be an original grammatogeny, for the signs do not denote syllables. Each hieroglyph denotes either one, two, or three consonants, or else is a true logogram representing only one word, or is a determinative. Thus is *p*, is *r*, and is *pr*; but is ambiguous. However, *pr* is 'house' and *pr* $^{r\text{LEGS}}$ is 'go forth' (presumably the two words had different vowels, but those are lost to us). How, then, did Egyptian writing come about? By "stimulus diffusion" from Sumerian (Daniels 2007b [pub. 2015]). Archeologists have demonstrated contacts between pre-Dynastic Egypt (late fourth millennium BCE) and Mesopotamia. An Egyptian could have seen, or heard of, this "writing" thing the Sumerians were doing, inquired, and learned that one made a mark for each separate word. In Sumerian, those words didn't change; but Egyptian accomplished inflecting by changing the vowels within words (like its distant relative Semitic), so what didn't change for the Egyptians was the consonants of the morphemes; so the signs, for them, represented only the consonants. The shapes of the Egyptian signs were already in existence, used on labels that have not yet been satisfactorily interpreted.[25] The first Egyptian scribe, then, put together an existing accounting(?) system with a vague idea of the workings of cuneiform orthography, and hieroglyphic writing came about.

The second accidental advance in writing yielded the abjad. In Egypt (probably) were speakers of a Semitic language (probably) who wanted to be able to compose prayers in their own language (probably). In the eighteenth century BCE or so – Egypt's Twelfth Dynasty – they didn't have the luxury of attending Egyptian scribal school for the years it took to master Egyptian writing; but someone did find out that what the hieroglyphs stood for was the consonants of the language. This innovator did not learn to associate the Egyptian sounds with the shapes, but instead used the initial sounds of Semitic words that named the items shown by the pictograms. Thus is not *r* as in Egyptian, but *p* as in Akkadian *pû*, Hebrew *pê* 'mouth' – the ancestor of פ and ڢ and ف and even ∏ and P.

Third, there was an encounter between a Greek and a Phoenician. Most likely it occurred in some mercantile center, probably in the eastern Mediterranean, where the two peoples interacted and traded.[26] The most likely time is ca. 800 BCE (Sass 2005). A Greek merchant saw a Phoenician merchant making records of sales, and wondered what was going on. The Phoenician – perhaps

a good friend – may have violated trade secrets principles by explaining the abjad: that each of the 22 letters represented the first sound of its name: *ʔalp*, *bayt*, *gaml*, etc., and that by sounding out the word, you could make a note of it. The Greek grasped the principle easily enough – but wasn't very good at Phoenician; hadn't learned to pronounce [ʔ] – or several other consonants that are found in Phoenician but not Greek. Thus what the Greek heard was *alp*, and also *e* (not *hê*), *ē* (not *ḥēt*), *i* (not *yôd*), *o* (not *ʕayn*), and *u* (not *wāw*). The Greek quite happily sounded out words and found both consonants and vowels that matched up with the beginnings of the letter names.[27]

2.3 *External characteristics*

All writing systems have properties relating to their physical existence that have no bearing on their function and so can vary widely. What is omitted in Section 1 is the change of shape of letters with time and space. On a concrete level, such changes reflect a constant struggle between economy of material and effort on the one hand, and ease of legibility on the other. More abstractly, letterforms are likely to conform to the visual esthetics prevailing more generally in the culture concerned.[28]

The shapes of characters can be influenced by the materials on which and with which they are written: ink or other pigmented liquid, applied with a flexible animal hair or vegetal fiber brush or a fairly stiff pen of reed, quill, or metal to a solid or flexible vegetal (papyrus, paper) or animal (skin, leather, parchment) surface; a rigid pointed burin incising a solid or palm-leaf surface; or a stylus on yielding clay. Any signs can be imitated by carving or molding all sorts of materials, and of course using the modern media of printing from movable type and viewing electronic fonts.

The elements of a script cannot be tossed randomly onto a surface: as the speech stream proceeds unidirectionally through time, so also does the writing stream progress linearly through space – though the continuing availability of the written text marks a profound difference between written and spoken language – and many orientations have been used: horizontal rows (almost always top to bottom) read left-to-right, right-to-left, or both directions in alternate lines; vertical columns (top to bottom), either right-to-left or left-to-right; even outward spirals.

A property possessed by many writing systems with a limited inventory of signs is a canonical order in which the signs are learned and which becomes an organizing principle for lists. Such orders may be arbitrary or motivated; and virtually the only motivated sign-order is the phonetic one established in India,[29] which places the vowels before the occlusives (back to front of mouth), followed by the continuants.

Countless explanations for the familiar *a, b, c, …* order, which is attested as far back as 1200 BCE, have been proposed (see Driver 1976: 179–85, 269–73 for summaries and refutations). What seems to me the most likely explanation for this order – as well as for the order Sequoyah himself used for his syllabary – is simply that it was the order in which the devisers happened to think of the sounds being memorialized. This accounts for occasional adjacencies like *wāw* 'mace' / *zayin* 'weapon', *yōd* 'hand' / *kap* 'palm of hand' (semantic association) and possibly *b g d* (phonetic association). The reason the word *abjad* exists (see Section 1.3) is that Arabic usually uses a motivated order derived from the earlier one: the letters with the same basic shape are brought together; but the reason the word *abugida* exists (see Section 1.4) is that since earliest times – again, attested ca. 1200 BCE – a different order has been used for the South Semitic abjads,[30] but the Hebrew order is known from the superscripts to the 8-verse acrostic (in Hebrew) portions of Psalm 119, which are transliterated in the Ge'ez Bible.

Letters added to an abjad or alphabet are usually ordered at the end, as happened with the local Greek alphabets. Sometimes letters are inserted according to graphic similarity, as in Arabic, and sometimes phonetic, as in Cyrillic. But even here, arbitrariness can prevail, as with Armenian.

For letters to be learned in an order, they need to have names, which may either be words in the language they record or reflect in arbitrary patterns their sound. We have seen how the names of letters relate to the pictograms underlying their shapes by the *acrophonic* principle (the initial sound of the name is the sound represented). Sometimes, as with Greek, the names are taken over along with the letters (losing the connection with the pictograms, which no longer existed anyway); sometimes new names are assigned (as with the North Germanic runes and the Russian alphabet).

3 Writing and Language

The theoretical aspects of writing systems presented here are grounded in a fundamental observation: writing is not like language, and it is not like language for biological reasons. The human language faculty evolved over some many generations, so that no human infant can avoid learning the language(s) of the environment. No child, however, can learn to read or write simply by watching other people read or write: explicit instruction is required. For writing is so recent (anthropologically speaking) that no special capacity for it can have evolved – especially since literate populations have not reproduced in preference to nonliterate ones![31] Moreover, there is no "critical period" for learning to read comparable to that for acquiring a native language; adult illiterates can become literate with no more difficulty than schoolchildren.

From this observation it follows that writing need not be structured or described in the same way as language, and in fact some language-derived analytical tools are not so well suited to writing. The linguistic terms phoneme, morpheme, and so on refer to an unconscious property of language and other realms of human behavior. Each item in a class of "-eme"-designated things is an abstraction, its identity defined by its contrasts with all the other items in that class, and comprising a group of instantiations of the thing (Headland, Pike, and Harris 1990). Every language includes a fairly small inventory of phonemes, and every morpheme is realized with phonemes; every stretch of speech is made up entirely of morphemes, which are made up entirely of phonemes. Here the unconscious-ness is important: since writing is not an unconscious, built-in feature of a mind (as language is), it cannot a priori be assumed to be analyzable in a parallel way. Rather, all writing systems were at some point consciously devised. The phonemic organization of various phenomena was recognized only a century or so ago, so it is not surprising that the designed writing of language differs in several ways from the evolved speaking of language.

- Writing systems, unlike languages, do not all operate in the same way. Different writing systems relate to the sound systems they record in fundamentally different ways (Section 1). These concern both the amount of speech each symbol represents, and the level of analysis the symbols embody.
- Despite American structuralist attempts to approach writing as a subsystem of language, writing systems do not work like linguistic systems; there is no "emic" level, and the popular term *grapheme* is misleading. For instance, many alphabets use a pairing of symbols – capitals and lowercase, majuscule and minuscule – that has no equivalent in sound systems, and no coherent definition of grapheme can be agreed on. Is it (like a phoneme) one of the set of elements comprising a writing system? Is it (like a tagmeme) a correlation of sound and symbol? Is it (like a morpheme) a minimal extent of something? The difficulty is that all these characterizations are reasonable for different writing systems, but no one characterization fits everything one might be tempted to call a grapheme. The upshot is that *grapheme* has become nothing more than a pre-theoretic, fancy, scientific-sounding word for "letter" or "character" and ought not to be part of technical discourse. ("Allograph," however, remains useful for conditioned variants of lettershapes.)

- Language is constantly changing, while writing generally obeys tradition and does not readily respond to changes. Simplification in some areas of language is accompanied by complication in other areas, as a language's overall "efficiency" tends to remain constant; but a script's efficiency – its "goodness of fit" to its language – is maximal when it is devised, and deteriorates thereafter.
- Writing systems can be altered by fiat; languages cannot.
- Probably most important, written language differs in significant ways from spoken language; the way most directly related to the physical existence of writing is the evanescence of speech versus the protracted availability of writing. The study of written language is a highly productive field separate from the study of writing systems (e.g., Olson and Torrance 2009).

Writing systems, then, must be investigated on their own terms. Their changes in appearance over the centuries – their "outer form" – have attracted the most study and are well documented (note 2), but what I find more interesting are questions of the origin of writing and the relation of the graphic shapes of script to the phonological shapes of language – their "inner form" (Coulmas 1996: 234).[32]

4 The Study of Writing

Writing is indispensable for civilization – under one definition of "civilization," perhaps the one that induces some to insist that any semasiographic system (Section 2) is "writing" – but entirely irrelevant for language. Most of the thousands of human languages were never written until recent years, and their speakers were none the worse for it. Their cultures were full and rich, lacking only accountancy and science. Everything else that is written need not be: poetry, narrative, and law, and their apotheosis, scripture, are all part of every oral culture. Only in a city[33] is the community so large that letters must be sent to communicate personal messages – and only when records of commerce can be kept can a city *be*. Cities are where production does not link directly with consumption: farmers and ranchers provide food, artisans provide goods, builders provide shelter, and administrators coordinate and tax their exchange. Without writing, there is no administration.

But cities characterize only a handful of human societies. The discovery that languages other than the classical ones were every bit as rich as Greek, Sanskrit, and Chinese – a discovery due largely to the investigation of African and Native American languages by scholars trained as Indo-Europeanists – led linguists to concentrate on unwritten languages and then to devalue the study of written records in favor of fieldwork. A reaction (associated with the "Toronto school" of literacy studies) to this view found some acceptance, which in its most extreme form (epitomized in Havelock 1986) claims that there was no true literacy before the Greek alphabet, even in the ancient Near East: that the alphabet itself is necessary for elevated linguistic expression. What this attitude reveals is little more than ignorance of both the literary record of nonalphabetic societies (which is generally known to the partisans only through translation) and of the poetic accomplishments of nonliterate societies (represented most familiarly, of course, in the supposed foundational work of western literacy, the Homeric epics).

A not dissimilar but differently motivated proposal suggests that the alphabet is the "best" sort of writing system (Diringer 1968) or, teleologically, the culmination of writing evolution, as encapsulated in Gelb's (1952: 201) elegant, but counterfactual, yet still widely cited, "Principle of Unidirectional Development," that "writing … must pass through the stages of logography, syllabography, and alphabetography in this, and no other, order." It was in order to make this formula come out right that Gelb had to call the Phoenician abjad a syllabary with indeterminate vowels and to essentially dismiss the abugidas of both India and Ethiopia; and of course only one alphabet ever evolved, and a general principle can hardly be derived from a single example.

The branch of scholarship that has been concerned with writing systems is *philology*, the study of texts in all their complexity.[34] It studies inscriptions (*epigraphy*), manuscripts (*paleography*), and documents (*diplomatics*); it deciphers scripts, and the philologist is often the partner of the archeologist as new texts are found.[35] The historical study of languages was *comparative philology*, when what was studied were ancient texts, before it was overtaken by *comparative linguistics*, which studies living languages. The two fields, when both are available, are complementary.

Is the study of writing – *grammatology*, as Gelb dubbed it (1952), or *graphonomy*, for Hockett (1951) – to be seen as a part of linguistics? The study of written language certainly is. But the fundamental difference between language and writing suggests that perhaps writing is outside the scope of linguistics, especially when linguistics is seen as a part of psychology. Perhaps graphonomy is truly a sister science under the umbrella of semiotics. Semiotic approaches to writing, however, have tended to slight philological concerns, to skip right over the details in favor of ungrounded theorizing. Perhaps as writing systems come back into the ken of linguists, the situation will improve.

NOTES

1 Sampson 1985; Coulmas 1989; DeFrancis 1989.

2 Taylor 1883; Cohen 1958; Février 1959; Friedrich 1966; Diringer 1968; Jensen 1969; and Senner 1989. On a smaller scale, but useful, are Nakanishi 1980 and especially Woodard 1996. Gnanadesikan 2009 is a superbly written presentation that does not always incorporate materials and developments since *WWS*.

3 A notion associated with the name of I. J. Gelb, but as he observes (1952: 284 n. 12), the term, and perhaps even the notion, was not his, but A. L. Kroeber's.

4 Note the title of Pike 1943, *Phonemics: A Technique for Reducing Languages to Writing*. It is increasingly recognized that the very notion of the phoneme as the fundamental unit of analysis of speech is an artifact of West European alphabetic literacy.

5 Orthographies adopting this approach, besides English, include Thai and Tibetan.

6 It's very unusual for an orthography to reflect *phonetic* rather than phonological reality: the two generally suspected examples are Masoretic Hebrew of the Jewish scriptures, and the Avestan alphabet of the Zoroastrian scriptures.

7 Two early north European scripts derived from the Latin: Germanic runes (1st–10th centuries CE) and Celtic ogham (5th–7th).

8 Transposing the characters of a script into a more familiar one is *transliteration*; laying out essentially all the linguistic information conveyed is *transcription*.

9 I gratefully acknowledge the assistance of John Huehnergard and Jo Ann Hackett.

10 Both scripts descend from the Aramaic rather than the Phoenician abjad, different though they appear. Early Hebrew inscriptions used a Phoenician abjad, but the familiar "Square Hebrew" or "Assyrian" script is associated with the "Babylonian Exile," during which generations of the soon-to-be-Jewish elite were trained in Aramaic. See Finkel 2014 for discussion of the evidence from the Mesopotamian side for this interaction.

11 The first word differs slightly: *bʾrīšīṯ*.

12 As with *abjad*, the consonants reflect the ancestral Semitic order; the vowels are in the traditional order of the "orders" (see below). Recently the letter-order of both consonants and vowels used in ancient India has been recognized (Salomon 2006), and if I were devising a term for the writing-system type today, I might choose *arepiconu*.

13 A frequently seen term is *alphasyllabary*. This was introduced (Bright 1992) about the same time as *abugida* (Daniels 1990), and the difference is summarized as "formal vs. functional" respectively in Daniels 1996, 4 n. *, and explicated in Bright 2000. Other existing terms – *neosyllabary*, *pseudo-alphabet*, *semi-syllabary*, *syllabically organized alphabet* – like *alphasyllabary* are unsatisfactory because they obscure the essential distinction from both syllabary and alphabet.

14 This order is often transliterated with ə, but this misleadingly suggests a reduced vowel, a shwa.

15 But not an alphasyllabary (Bright 2000) – the vowels other than /a/ are given by full-size letters.

16 *Semiotic*: conveying meaning. *Semasiographic*: visually recording meaning.

17 Other components of this script are ð, ɪ, ß, and þ, all of which occur in one or another European alphabet.
18 I use "-gram" rather than "-graph" in these words solely on the model of *telegram* vs. *telegraph*, where, of the two, the former is the information-bearing item.
19 Sumerian writing is incised on moist clay with a rectangular-cross-section stylus. The pictographic period didn't last very long; the images were soon stylized into combinations of straight stylus impressions made by touching an edge or corner to the clay surface, and over the centuries the characters simplified. The three shown in the text ended up as *ti*, *en*, and *gub*.
20 The monosyllabic hypothesis adds some weight to the supposition that the language underlying the Indus Valley script, known from hundreds of short inscriptions, may be a form of Dravidian (and not Munda or Indo-European).
21 A considerable amount of morphophonemic alternation has now been identified in Sumerian. Is it possible that the dearth of stable forms of affixes prevented early scribes from recognizing that they could be notated, either with an abstracted base form, or with a plethora of different symbols for the different pronunciations each affix might exhibit?
22 An attempt to revive the vertical, connected Mongolian script with the ouster of Communism from Mongolia in 1989, to replace the Cyrillic alphabet imposed by the Russians, met little success.
23 It is not an alphasyllabary because the other vowels are given by full letters rather than "diacritics" (Bright 2000).
24 In previous accounts, I didn't count this as a distinct mode.
25 They were discovered only late in the twentieth century. A few examples were found more than a century ago but were simply reported and filed away.
26 Jeffery (1990) suggested Al-Mina, quite possibly because at the time she was first writing, that was the only site that had been sufficiently investigated to demonstrate the presence of both Greeks and Phoenicians.
27 It has been suggested that the Greek alphabet has an Aramaic rather than a Phoenician origin, because several of the letter names end with *a* (*alpha*, *beta*, etc.), reflecting the Aramaic "emphatic state" of the names, *ʔalpā*, *bētā*, etc. But if Greek had encountered vowel letters in an orthography, it's unlikely that they would have been assigned a very different sort of use.
28 Fuller discussion of the subjects covered in Section 2.3 may be found in the first edition of this chapter (http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1405102527.html)
29 The exceptions are Japanese, where the classical arrangement of the kana sets spells out a poem ("*Iroha*"), and Javanese, where the 20 abugidic letters spell out a sentence summarizing an etiological tale.
30 Kammerzell 2001 offers an intriguing but unlikely reconstruction of an Egyptian order that could underlie both the Northwest and the South Semitic orders.
31 This is not to say, however, that there are not uniformities in the way writing and reading are processed in the human brain across individuals, though investigations involving non-alphabetic scripts have barely begun (Dehaene, Cohen, Morais, and Kolinsky 2015).
32 Coulmas 1996 contains numerous insightful articles on societal aspects of writing (see also Coulmas 2013). Unfortunately the work is arguably unreliable as to factual matters.
33 Latin *civis*, thus etymologically the basis of *civi*lization.
34 Summaries of these fields may be found in the first edition of this chapter (http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1405102527.html), and in Daniels 2005.
35 The standard work on decipherment remains Pope 1999, supplemented by Daniels 1994, with the more reader-friendly Robinson 2002. Scripts deciphered or reinterpreted since those books were written include Carian (Adiego 2007), Udi (Gippert et al. 2009–10), and Meroitic (Rilly and de Voogt 2012).

REFERENCES

Adiego, I-J. (2007). *The Carian Language*. Handbuch der Orientalistik 1.86. Leiden: Brill.

Baines, J., Bennet, J., and Houston, S. D. (eds) (2008). *The Disappearance of Writing Systems: Perspectives on Literacy and Communication*. London: Equinox.

Baxter, W. H. and Sagart, L. (2014). *Old Chinese: A New Reconstruction*. Oxford: Oxford University Press.

Brengelman, F. H. (1980). Orthoepists, printers, and the rationalization of English spelling. *Journal of English and Germanic Philology*, 79, 332–54.

Bright, W. (ed.) (1992). *International Encyclopedia of Linguistics*. New York: Oxford University Press.

Bright, W. (2000). A matter of typology: alphasyllabaries and abugidas. *Studies in the Linguistic Sciences* (Urbana), 30, 63–71.

Burnaby, B. (ed.) (1985). *Promoting Native Writing Systems in Canada*. Ontario Institute for Studies in Education Occasional Papers 24. Toronto: OISE Press.

Cohen, M. (1958). *La grande invention de l'écriture et son évolution*. Paris: Imprimerie Nationale.

Coulmas, F. (1989). *The Writing Systems of the World*. Oxford: Blackwell.

Coulmas, F. (1996). *The Blackwell Encyclopedia of Writing Systems*. Oxford: Blackwell.

Coulmas, F. (2005). *Writing Sytems*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Coulmas, F. (2013). *Writing and Society: An Introduction*. Key Topics in Sociolinguistics. Cambridge: Cambridge University Press.

Daniels, P. T. (1990). Fundamentals of grammatology. *Journal of the American Oriental Society*, 110, 727–31.

Daniels, P. T. (1992a). Contacts between Semitic and Indic scripts. In A. Harrak (ed.), *Contacts between Cultures: Selected Papers from the 33rd International Congress of Asian and North African Studies, Toronto, August 15–25, 1990*, vol. 1, *West Asia and North Africa* (pp. 146–52). Lewiston, NY: Edwin Mellen.

Daniels, P. T. (1992b). The syllabic origin of writing and the segmental origin of the alphabet. In P. Downing, S. D. Lima, and M. Noonan (eds), *The Linguistics of Literacy* (pp. 83–110). Amsterdam/Philadelphia, PA: John Benjamins.

Daniels, P. T. (1994). Edward Hincks's decipherment of Mesopotamian cuneiform. In K. J. Cathcart (ed.), *The Edward Hincks Bicentenary Lectures* (pp. 30–57). Dublin: University College Dublin, Department of Near Eastern Studies.

Daniels, P. T. (1996). The study of writing systems. In P. T. Daniels and W. Bright (eds), *The World's Writing Systems* (pp. 3–17). New York: Oxford University Press.

Daniels, P. T. (2000). Review article on popular books on writing. *Sino-Platonic Papers*, 98, 47–57. http://sino-platonic.org/complete/spp098_book_reviews.pdf

Daniels, P. T. (2005). Philology. In P. Strazny (ed.), *Encyclopedia of Linguistics* (pp. 824–26). New York: Fitzroy Dearborn.

Daniels, P. T. (2007a). *Littera ex occidente*: toward a functional history of writing. In C. L. Miller (ed.), *Studies in Semitic and Afroasiatic Linguistics Presented to Gene B. Gragg* (pp. 53–68). Studies in Ancient and Oriental Civilization 60. Chicago, IL: Oriental Institute of the University of Chicago. (http://oi.uchicago.edu/pdf/saoc60.pdf).

Daniels, P. T. (2007b [pub. 2015]). Three models of script transfer. *Word*, 57(3), 371–78.

Daniels, P. T. (2009). Grammatology. In D. R. Olson and N. Torrance (eds), *The Cambridge Handbook of Literacy* (pp. 25–45). Cambridge: Cambridge University Press.

Daniels, P. T. (2010). Writing in the world and linguistics. *BLS*, 36, 61-84. http://escholarship.org/uc/item/7vv3w7p8

Daniels, P. T. (2013). The history of writing as a history of linguistics. In K. Allan (ed.), *Oxford Handbook of the History of Linguistics* (pp. 53–69). Oxford: Oxford University Press.

Daniels, P. T. (in press). *An Exploration of Writing*. Sheffield: Equinox.

Daniels, P. T. and Bright, W. (eds) (1996). *The World's Writing Systems*. New York: Oxford University Press.

DeFrancis, J. (1989). *Visible Speech: The Diverse Oneness of Writing Systems*. Honolulu: University of Hawaii Press.

Dehaene, S., Cohen, L., Morais, J., and Kolinsky, R. (2015). "Illiterate to literate: behavioural and cerebral changes induced by reading acquisition." *Nature Reviews: Neuroscience*, 16, 234–44. DOI 10.1038/nrn3924.

Diringer, D. (1968). *The Alphabet: A Key to the History of Mankind*, edited by R. Regensburger, 3rd edn. New York: Funk & Wagnall's. (1st edn, 1948.)

Driver, G. R. (1976). *Semitic Writing*, edited by S. A. Hopkins, 3rd edn. Schweich Lectures, 1944. London: Oxford University Press for the British Academy.

Février, J.-G. (1959). *Histoire de l'écriture*, 2nd edn. Paris: Payot.

Finkel, I. 2014. *The Ark before Noah: Decoding the story of the Flood*. London: Hodder & Stoughton.

Foreman, G. (1938). *Sequoyah*. Norman, OK: University of Oklahoma Press.

Frellesvig, B. (2010). *A History of the Japanese Language*. Cambridge: Cambridge University Press.

Friedrich, J. (1966). *Geschichte der Schrift*. Wiesbaden: Harrassowitz.

Gelb, I. J. (1952). *A Study of Writing*. Chicago: University of Chicago Press. (Reissued with additions and corrections, 1963.)

Gippert, J., Schulze, W., Aleksidzé, Z., and Mahé, J.-P. (2009–10). *The Caucasian Albanian Palimpsests of Mount Sinai*. 3 vols. Turnhout: Brepols.

Gnanadesikan, A. E. (2009). *The Writing Revolution: Cuneiform to the Internet*. Malden, MA: Wiley-Blackwell.

Havelock, E. (1986). *The Muse Learns to Write: Reflections on Orality and Literacy from Antiquity to the Present*. New Haven, CT: Yale University Press.

Headland, T. N., Pike, K. L., and Harris, M. (eds) (1990). *Emics and Etics: The Insider/Outsider Debate*. Frontiers of Anthropology 7. Newbury Park, CA: Sage.

Hockett, C. F. (1951). Review of *Nationalism and Language Reform in China*, by J. DeFrancis. *Language*, 27(3), 439–45.

Hockett, C. F. (1997). Review of *The World's Writing Systems*, ed. P. T. Daniels and W. Bright. *Language*, 73, 379–85.

Houston, S. D. (ed.) (2004). *First Writing: Script Invention as History and Process*. Cambridge: Cambridge University Press.

Houston, S. D. (ed.) (2012). *The Shape of Script: How and Why Writing Systems Change*. Santa Fe, NM: School for Advanced Research Press.

Jeffery, L. H. (1990). *The Local Scripts of Archaic Greece*, edited by A. W. Johnston, 2nd edn. Oxford: Clarendon.

Jensen, H. (1969). *Sign, Symbol and Script*, translated by G. Unwin. New York: Putnam's.

Kammerzell, F. (2001). Die Entstehung der Alphabetreihe: Zum ägyptischen Ursprung der semitischen und westlichen Schriften. In D. Borchers, F. Kammerzell, and S. Weninger (eds), *Hieroglyphen Alphabete Schriftreformen: Studien zu Multiliteralismus, Schriftwechsel und Orthographieneuregelungen* (pp. 117–58). *Lingua Aegyptia* Studia monographica 3. Göttingen: Seminar für Ägyptologie und Koptologie.

Kara, G. (2005). *Books of the Mongolian Nomads: More than Eight Centuries of Writing Mongolian*, translated by J. R. Krueger. 1st English edn, revised and expanded by the author. Indiana University Ural and Altaic Series 171. Bloomington, IN: Indiana University Research Institute for Inner Asian Studies. (Russian original, 1972.)

Kim-Renaud, Y.-K. (ed.) (1997). *The Korean Alphabet*. Honolulu: University of Hawaii Press.

Klugkist, A. C. (1982). Midden-aramese schriften in Syrië, Mesopotamië, Perzië en aangrenzende gebieden. PhD dissertation, Rijksuniversiteit te Groningen.

Ledyard, G. (1997). The international linguistic background of the correct sounds for the instruction of the people. In Y.-K. Kim-Renaud (ed.) *The Korean Alphabet* (pp. 31–87). Honolulu: University of Hawaii Press.

Mafundikwa, S. (2004). *Afrikan Alphabets: The Story of Writing in Afrika*. West New York, NJ: Mark Batty.

Michalowski, P. (2004). Sumerian. In R. D. Woodard (ed.) *Encyclopedia of the World's Ancient Languages* (pp. 19–59). Cambridge: Cambridge University Press.

Morag, S. (1961). *The Vocalization Systems of Arabic, Hebrew, and Aramaic*. Janua Linguarum Series Minor 13. The Hague: Mouton.

Mora-Marín, D. (2003). The origin of Mayan syllabograms and orthographic conventions. *Written Language and Literacy*, 6, 193–238.

Mora-Marín, D. (2010). Consonant deletion, obligatory synharmony, typical suffixing: an explanation of spelling practices in Mayan writing. *Written Language and Literacy*, 13, 118–79.

Mumin, M. and Versteegh, K. (eds) (2014). *The Arabic Script in Africa: Studies in the Use of a Writing System*. Studies in Semitic Languages and Linguistics, 71. Leiden: Brill.

Nag, S. (2014). Akshara–phonology mappings: the common yet uncommon case of the consonant cluster. *Writing Systems Research*, 6, 105–19.

Nakanishi, A. (1980). *Writing Systems of the World: Alphabets, Syllabaries, Pictograms*. Rutland, VT: Tuttle.

Naveh, J. (1987). *Early History of the Alphabet*, 2nd edn. Jerusalem: Magnes.

O'Connor, M. (1983). Writing systems, native speaker analyses, and the earliest stages of Northwest Semitic orthography. In C. L. Meyers and M. O'Connor (eds), *The Word of the Lord Shall Go Forth: Essays in Honor of David Noel Freedman in Celebration of His Sixtieth Birthday* (pp. 439–65). Winona Lake, IN: Eisenbrauns.

Olson, D. R. and Torrance, N. (eds) (2009). *The Cambridge Handbook of Literacy*. Cambridge: Cambridge University Press.

Pike, K. L. (1943). *Phonemics: A Technique for Reducing Languages to Writing*. Glendale, CA: Summer Institute of Linguistics.

Pope, M. (1999). *The Story of Decipherment: From Egyptian Hieroglyphs to Maya Script*, 2nd edn. New York: Thames & Hudson. (1st edn, as *The Story of Archaeological Decipherment: From Egyptian Hieroglyphics to Linear B*, 1975.)

Qiu, X. (2000). *Chinese Writing*, translated by G. L. Mattos and J. Norman. *Early China* Special Monograph 4. Berkeley: Society for the Study of Early China & Institute of Early Asian Studies, University of California, Berkeley. (Chinese originals, 1988 [Beijing] and 1994 [Taipei].)

Rilly, C. and de Voogt, A. (2012). *The Meroitic Language and Writing System*. Cambridge: Cambridge University Press.

Robinson, A. (2002). *Lost Languages*. New York: McGraw Hill.

Salomon, R. G. (2006). Kharoṣṭhī syllables used as location markers in Gandhāran stūpa architecture. In Pierfrancesco Callieri (ed.) *Architetti, capomastri, artigiani: L'organizzazione dei cantieri e della produzione artistica nell'Asia ellenistica, Studi offerti*

a Domenico Faccenna nel suo ottantesimo complean (pp. 181–224). Serie Orientale Roma 100. Rome: Istituto Italiano per l'Africa e l'Oriente.

Sampson, G. (1985). *Writing Systems*. London; Stanford, CA: Hutchinson; Stanford University Press. (Corrected pbk. reprint, London, 1987.)

Sass, B. (2005). *The Alphabet at the Turn of the Millennium*. Tel Aviv Occasional Publications 4. Tel Aviv: Tel Aviv University.

Scancarelli, J. (1992). Aspiration and Cherokee orthographies. In P. Downing, S. D. Lima, and M. Noonan (eds), *The Linguistics of Literacy* (pp. 135–52). Typological Studies in Language 21. Amsterdam/Philadelphia, PA: John Benjamins.

Schmitt, A. (1980). *Entstehung und Entwicklung von Schriften*, edited by C. Haeber. Cologne: Böhlau.

Scribner, S. and Cole, M. (1981). *The Psychology of Literacy*. Cambridge, MA: Harvard University Press.

Senner, W. (ed.) (1989). *The Origins of Writing*. Lincoln, NE: University of Nebraska Press.

Share, D. L. and Daniels, P. T. (2015). Aksharas, alphasyllabaries, abugidas, alphabets, and "orthographic depth": reflections on Rimzhim, Fowler, and Katz (2014). *Writing Systems Research*, 8 (2016): 17–31. DOI 10.1080/17586801.2015.1016395.

Sims-Williams, N. (1981). The Sogdian sound-system and the origin of the Uyghur script. *Journal Asiatique*, 347–59.

Taylor, I. (1883). *The Alphabet: An Account of the Origin and Development of Letters*. London: Kegan Paul, Trench. (2nd edn, London: Edward Arnold, 1899.)

Tuchscherer, K. and Hair, P. E. H. (2002). Cherokee and West Africa: examining the origins of the West African script. *History in Africa*, 29, 427–86.

van Schaik, S. (2011). A new look at the Tibetan invention of writing. In Y. Imaeda, M. T. Kapstein, and T. Takeushi (eds), *New Studies of the Old Tibetan Documents: Philology, History and Religion* (pp. 45–96). Old Tibetan documents online monograph series 3. Tokyo: Tokyo University of Foreign Studies, Research Institute for Languages and Cultures of Asia and Africa. Retrieved January 27, 2014 (https://otdo.aa.tufs.ac.jp).

Walker, W. and Sarbaugh, J. (1993). The early history of the Cherokee syllabary. *Ethnohistory*, 40(1), 70–94.

Webster, N. (1806). *A Compendious Dictionary of the English Language*. New Haven: Sidney's Press. (Reprint, edited by Philip B. Gove. N.p: Crown Publishers, 1970.)

Woodard, R. D. (1996). Writing systems. In *The Atlas of Languages: The Origin and Development of Languages throughout the World*, 162–209. London; New York: Quarto; Facts on File.

Wyrod, C. (2008). A social orthography of identity: the N'ko literacy movement in West Africa. *International Journal of the Sociology of Language*, 192, 27–44.

Zimmer, B. (2010). On language: is this word really pronounced 'fish'? *New York Times Magazine*, June 27, 14. (http://www.nytimes.com/2010/06/27/magazine/27FOB-onlanguage-t.html).

Part II Theoretical Bases

6 The History of Linguistics

Approaches to Linguistics

LYLE CAMPBELL

1 Introduction

Many "histories" of linguistics have been written over the last two hundred years, and since the 1970s linguistic historiography has become a specialized subfield, with conferences, professional organizations, and journals of its own. Works on the history of linguistics often had such goals as defending a particular school of thought, promoting nationalism in various countries, or focusing on a particular topic or subfield, for example on the history of phonetics. Histories of linguistics often copied from one another, uncritically repeating popular but inaccurate interpretations; they also tended to see the history of linguistics as continuous and cumulative, though more recently some scholars have stressed the discontinuities. Also, the history of linguistics has had to deal with the vastness of the subject matter. Early developments in linguistics were considered part of philosophy, rhetoric, logic, psychology, biology, pedagogy, poetics, and religion, making it difficult to separate the history of linguistics from intellectual history in general, and, as a consequence, work in the history of linguistics has contributed also to the general history of ideas. Still, scholars have often interpreted the past based on modern linguistic thought, distorting how matters were seen in their own time. It is not possible to understand developments in linguistics without taking into account their historical and cultural contexts. In this chapter I attempt to present an overview of the major developments in the history of linguistics, avoiding these difficulties as far as possible.

2 Grammatical Traditions

A number of linguistic traditions arose in antiquity, most as responses to linguistic change and religious concerns. For example, in the case of *the Old-Babylonian tradition*, when the first linguistic texts were composed, Sumerian, which was the language of religious and legal texts, was being replaced by Akkadian. This grammatical tradition emerged by about 1900 BCE and lasted 2,500 years, so that Sumerian could be learned and these texts could continue to be read. Most of the texts were administrative lists: inventories, receipts, and rosters. Some early texts for use in the scribal school were inventories (lists) of Sumerian nouns and their Akkadian equivalents.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

From this, grammatical analysis evolved in the sixth and fifth centuries BCE; different forms of the same word, especially of verbs, were listed in a way that represented grammatical paradigms and matched them between the two languages (Gragg 1995; Hovdhaugen 1982).

Language change also stimulated the *Hindu tradition*. The Vedas, the oldest of the Sanskrit memorized religious texts, date from ca. 1200 BCE. Sanskrit, the sacred language, was changing, but ritual required exact verbal performance. Rules of grammar were set out for learning and understanding the archaic language. Pāṇini's (ca. 500 BCE) description (which also contains rules formulated by his predecessors, in a tradition from the tenth to the seventh centuries BCE) originated in comparisons between versions called *padapāta* (word-for-word recitation) and *samhitapāta* (continuous recitation, of divine origin, unalterable) of the same Vedic texts. The grammatical rules were devised for this comparison and for checking textual accuracy, and technical methods of grammatical description were developed in connection with the formulation of these rules. In addition to Pāṇini, Kātyāyana's rules of interpretation (ca. 300 BCE) and Patañjali's commentary (ca. 150 BCE) are important in this tradition. Grammar was considered the most scientific of the sciences in India, and the scholars in other areas aspired to the ideal embodied in the Hindu grammatical tradition (Staal 1974).

The *Greek grammatical tradition*, which also owes its origin to language change, was developed originally by schoolmasters, though it is known only from later writings of philosophers. Homer's works (ca. 850 BCE) were basic in early Greek education, but the Greek of the fifth to the third centuries BCE had changed so much that explanations of Homer's language were important in the school curriculum. Observations taken from earlier school grammar are found in works of Plato, Aristotle, and the Stoics (Hovdhaugen 1982: 46). Themes important in the ancient Greek tradition have persisted throughout the history of linguistics, such as the origin of language, parts of speech (grammatical categories), and the relation between language and thought, to mention just a few. A persistent controversy was whether "nature" or "convention" accounted for the relationship between words and their meaning, and this had implications for the history of language and for the origin of words. Earlier opinions on the matter are contrasted in Plato's (427–347 BCE) *Cratylus*. At issue was whether language originated in "nature" (*phúsis*), with the first words supposedly imitating the things that they name, or in "convention" (*nómos* or *thésis*), that is, in usage or naming, whether of human or divine invention, or in a synthesis of the two. Aristotle (384–322 BCE) in *De interpretatione* favored convention over nature; the Stoics held that language originated in nature.

For the Greeks, morphology (word structure) was mostly a historical matter, about the creation of the structure of words (part of "etymology"). Syntax was not described directly, but aspects of syntax were treated in rhetoric and logic. With respect to parts of speech, we see in Plato's division of the sentence into *ónoma* ("name") and *rħêma* ("utterance") an example where the interpretation of the past has been based too much on present understanding. Plato's terms are at times equated with the modern categories "noun" and "verb," respectively, but they equally had shades of "subject" and "predicate," and "topic" and "comment," or even entity and relation. The parts of speech (grammatical categories) as understood in traditional grammar developed more fully with the Stoics and others (Hovdhaugen 1982: 41, 48).

Roman linguistics continued Greek themes. Aelius Donatus' (fourth century CE) *Ars minor* and *Ars major* and Priscian's (sixth century CE) *Institutiones grammaticae* (18 volumes) became exceedingly important in the middle ages. Except for Varro (116–27 BCE) and Priscian, Roman grammarians also did not treat syntax (only parts of speech); rather, morphology dominated in an approach focused on noun declensions and verb conjugations (Hovdhaugen 1982: 87).

The Arabic grammatical tradition had roots in the Greek grammatical traditions, especially following Aristotle. For Arabic grammarians, the Arabic language was sacred and immutable as enshrined in the Qur'ān, and they were concerned with explaining why Arabic was perfect. For example, the system of inflectional endings was believed to be proof of the symmetry and logicalness of the language. The major impetus for grammatical study came from linguistic change

and the desire to preserve the integrity of the holy language of the Qur'ān. While no change was acknowledged in formal Arabic after the eighth century, the realization that the spoken Arabic of the eighth and ninth centuries was changing stimulated the development of Arabic grammatical study. Abū'l-Aswad ad-Du'alī (died ca. 688) is reputed to be the inventor of this grammatical tradition, which commenced seriously in the writings of al-Khalīl (died 791) and Sībawayhi (a Persian; died 804) (Owens 1988). *The Hebrew linguistic tradition* began with concern for establishing the correct Hebrew text of the Old Testament. Hebrew grammarians borrowed descriptive methods wholesale from the Arabic linguistic tradition and developed a system of analysis for the morphology (analysis of words into their meaningful parts). Between 900 and 1550, 91 authors composed 145 works on grammar that we know of. Saadya ben Joseph al-Fayyūmī (a.k.a. Saadya Gaon) (882–942) is generally held to be the first to produce a Hebrew grammar and dictionary (Téné 1995: 22). Ibn Janāḥ of Cordoba's *Kitāb al-Luma'*, written in Judeo-Arabic, was the first complete description of Hebrew. For Ibn Janāḥ (born 980 CE), Hebrew, Arabic, and all other languages had three parts of speech: noun, verb, and particles (as in the Arabic tradition, inherited from Aristotle). The tradition reached its peak in David Qimḥi's (ca. 1235) grammar, *Sepher mikhlol*, whose main features were analysis of verbal forms with a set of affixes and roots. This kind of analysis came to have a strong impact on European linguistics. Johannes Reuchlin's (1506) comprehensive *De rudimentis Hebraicis* introduced the Hebrew method of morphological analysis in Europe, and Theodor Bibliander (1548) recommended this analysis of words into roots and affixes for the study of all languages. He thought languages described in the Hebrew manner would be "in conformity with nature" and could therefore be meaningfully compared (Percival 1986).

Early Christian writers returned to the philosophical themes of Aristotle and the Stoics. Classical Latin grammars, mainly Donatus' *Ars minor*, were adapted to church education. Teachings of Roman grammarians were mixed with folk views in a Christian frame. In the seventh and eighth centuries, Donatus predominated, though ca. 830 Priscian's *Institutiones* replaced Donatus as the basic grammar, resulting in a new tradition of commentaries, the first steps towards the shift of interest in the eleventh and twelfth centuries which gave rise to the theory-oriented speculative grammar of the thirteenth and fourteenth centuries. The origin of languages was also of natural interest to the multilingual early Christian world, with notions of Babel and of taking the "word" to the nations of the earth (Hovdhaugen 1982: 109). In this environment, the hypothesis that Hebrew was the original language from which all others sprang became predominant.

3 The Rise of Universal Grammar

Around 1000 CE, a shift began in which logic came to dominate linguistic thought. Prior to 1100, most scholars adhered faithfully to Donatus and Priscian; from the twelfth century onward there was a return to dialectics. The recovery through Arabic scholarship of Aristotle's lost writings was an important factor, and Arabic commentators were quoted amply. Grammarians followed Aristotle's view that scientific knowledge is universal or general and applies to all subject matter, including grammar, hence universal grammar. Semantic analysis (or logical theory) came to dominate Europe for the next four centuries. Pierre Abailard's (Abelard's) (1079–1142) *Dialectica* (1970 [ca. 1130]) systematized logic as expressed through the structure of ordinary language, building on Aristotle and placing logic at the highest level of contemporary science. Robert Kilwardby (died 1279) insisted on the universal nature of grammar, a concept more fully developed by Roger Bacon (1214–1294), both Englishmen who taught in Paris. Bacon is famous for his statement that "grammar is substantially one and the same in all languages, although it may vary accidentally" (Bursill-Hall 1995: 131).

"*Speculative grammar*" developed, with concern for the notion of *modi significandi* "ways of signifying." Some 30 authors, called *Modistae*, most connected with the University of Paris, integrated Donatus and Priscian into scholastic philosophy (1200–1350), that is, the integration of

Aristotelian philosophy into Catholic theology. According to the *Modistae*, the grammarian's job was to explain how the intellect had created a system of grammar; in language the grammarian expressed understanding of the world and its contents through the modes of signifying (Bursill-Hall 1995: 132). Such a grammatical system had to mirror reality as grasped by understanding; that is, grammar was ultimately underwritten by the very structure of the universe (Breva-Claramonte 1983: 47). The *Modistae* compiled lists of modes of signifying for Donatus' and Priscian's parts of speech, distinguishing essential modes (the same in all languages) from accidental ones. For example, "predication" (verb) was essential to communication, but "tense" was accidental, since its function could be signified by something else, for example by temporal adverbs. "Noun" was the most essential (echoing Aristotle).

In the fourteenth century, teaching grammars began to compete with the scholastic commentaries, and the Modistic approach faded; however, there was a revival of philosophical grammar in the sixteenth century, begun with Julius Caesar Scaliger's (l'Escale) (1484–1558) *De causis linguae latinae* (1540). For Scaliger, grammar was part of philosophy, including the causation or creation of language from nature (hence the *de causis* in his title) (Breva-Claramonte 1983: 62). Francisco Sánchez (Sanctius) de las Brozas (1523–1601) in *Minerva, seu de causis linguae latinae* (1587) attempted to reconcile Plato and Aristotle by explaining that the "convention" favored by Aristotle was "reasoned," and, since reasoning is universal, God-given, it comes from "nature," which is what Sanctius believed Plato to have favored. Thus Sanctius' philosophy of language was "a rational discovery of the underlying 'perfection' or logic of language from which actual speech is derived" (Breva-Claramonte 1983: 15). Sanctius' universal grammar, in turn, influenced Arnauld and Lancelot's *Grammaire générale et raisonnée de Port Royal* (1660), and James Harris's (1709–1780) *Hermes* (1751), seminal in universal grammar theory.

In medieval manuscripts, the inflectional paradigms of Latin were explicated or annotated with forms from the vernacular languages. This pedagogical practice was combined in the seventeenth century with the revival of scholastic logical grammar in the *Grammaire générale et raisonnée de Port Royal* (Arnauld and Lancelot 1660). Following René Descartes (1596–1650), with human understanding taken to be the same for all people, scholars held the basic forms of thought to be the basis of every grammar; the particular grammatical systems of existing languages were merely approximations of the universal ideal, partly corrupted by neglect in usage. The principal concern was with the manifestation of universal semantic concepts in individual languages. In the seventeenth century, language studies came to be based on new theories of cognition and the philosophy of language, in particular on John Locke's (1632–1704) *Essay Concerning Human Understanding* (1690).

4 The Rise of the Comparative Method

Through voyages, conquests, trading, and colonialization from the sixteenth century onward, Europe became acquainted with a wide variety of languages. Information on languages from Africa, Asia, and America became available in the form of word lists, grammars, dictionaries, and religious texts, and attempts at classifying these languages followed. Historical linguistic interests had a background in the Greek tradition's nature-versus-convention debate about language origins and its interest in etymology, as well as in the biblically based notion of Hebrew as the original language (*Lingua Adamica, Lingua Paradisiaca*) from which all others were assumed to descend after the confounding of tongues at Babel. From the catalogue of languages and peoples in Genesis came the tradition of *Sprachlisten*, "inventories of known languages of the world successively fitted into the Biblical ('Mosaic') framework, usually placing Hebrew at the head, between the third and seventeenth centuries" (Robins 1990: 86; Borst 1959).

Large-scale word collections for language comparisons were a notable feature of the centuries after the Renaissance. Some landmarks were Konrad Gesner 1555, Gottfried Wilhem Leibniz

1717, Johan Christoph Adelung 1782, 1806, Lorenzo Hervás y Panduro 1784, 1800, Peter Simon Pallas 1786, among others. These played an important role in the development of comparative linguistics.

The development of comparative grammar is subject to interpretation, explaining why each of the following at one time or another has been considered the "father" of comparative linguistics: Giraldus Cambrensis 1194, Dante 1305, J. J. Scaliger 1610 [1599], Georg Stiernhielm 1671, Andreas Jäger 1686, Ludolf 1702, Adriaan Relander [Hadrianus Relandus] 1706, Edward Lhuyd 1707, Philip Johan Tabbert von Strahlenberg 1730, Johan Ihre 1769, Jo[h]annis [János] Sajnovics 1770, Sir William Jones 1798, Christian Kraus 1787, Sámuel Gyarmathi 1799, Franz Bopp 1816, 1833, Ramus Rask 1818, and Jacob Grimm 1818, among others. Hoenigswald's summary of the points upon which seventeenth- and eighteenth-century scholars agreed concerning criteria for establishing language families is telling:

> First, . . . there was "the concept of a no longer spoken parent language which in turn produced the major linguistic groups of Asia and Europe". Then there was . . . "a Scaliger concept of the development of languages into dialects and of dialects into new independent languages". Third came "certain minimum standards for determining what words are borrowed and what words are ancestral in a language", and, fourth, "an insistence that not a few random items, but a large number of words from the basic vocabulary should form the basis of comparison" . . . fifth, the doctrine that "grammar" is even more important than words; sixth, the idea that for an etymology to be valid the differences in sound – or in "letters" – must recur, under a principle sometimes referred to as "analogia". (1990: 119–20)

From the fifteenth century onward, etymology had been shifting away from its sense in classical antiquity of unfolding the true meaning of words toward a historical search for earlier stages in languages and the origin of words (Robins 1990: 86). Etymology thus became important in attempts to establish linguistic relationships. The *Dutch etymologists*, such as Scrieckius 1614, de Laet 1643, and ten Kate 1710, had a lasting impact. Their analysis of words into roots and affixes (prefixes and suffixes), which was inspired by the Hebrew grammatical tradition, became fundamental to the comparative method. They utilized three principal criteria for establishing family relationships which were to become standard: basic vocabulary, sound correspondences, and grammatical agreements.

4.1 The Scythian hypothesis and the notion of Indo-European

Eventually, comparative linguistics came to have Indo-European languages as its main concern. Early recognition of the family relationship among Indo-European languages is connected intimately with the "Scythian hypothesis." The *Scythae* of Classical writers (Herodotus, Strabo, Justin, etc.) were a nation on a sea in the north in extreme antiquity. Josephus and early Christian writers took them to be the descendants of Japheth (son of Noah), the assumed father of Europe (Droixhe 1984: 5), and the Scythian linguistic hypothesis emerged from these notions. Various proposals attempted to identify Scythians with different language groups of Europe and Asia, but proposed Indo-European associations came to dominate. With Johannes Goropius Becanus' (Jan van Gorp van Hilvarenbeek's) (1518–1572) (1569) emphasis on "Scythian," recognition of Indo-European as a language family began. Raphelengius (Ravlenghien) reported correspondences between Persian and Germanic languages. Marcus Zuerius Boxhorn(ius) (1602–1653) relied both on matches in words and on grammatical similarities to prove "that these people all learned their language from one same mother" (Muller 1986: 10). Others also advanced the Scythian hypothesis: Claudius Salmasius (Claude Saumaise) (1588–1653) (1643), Georg Stiernhielm (1598–1672) (1671), Andreas Jäger (1660–1730) (1686), Leibniz (1646–1716), and so on. So well known was the Scythian hypothesis that in 1733 Theodor Walter (1699–1741), a missionary in Malabar,

"recognized similarities between Sanskrit, Greek, and Persian numerals and explained these with . . . Scythian theory" (Fellman 1975: 38).

4.2 Sir William Jones

The most repeated passage in linguistic history is Sir William Jones's (1746–1794) statement in 1786:

> The *Sanscrit* language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either; yet bearing to both of them a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologer could examine them all three without believing them to have sprung from *some common source*, which, perhaps, no longer exists. There is a similar reason, though not quite so forcible, for supposing that both the *Gothic* and *Celtick*, though blended with a very different idiom, had the same origin with the *Sanscrit*; and the old *Persian* might be added to the same family, if this were the place for discussing any question concerning the antiquities of *Persia*. (Jones 1798: 422–3)

Based on this, Jones is usually credited with founding comparative linguistics and discovering the relationship among Indo-European languages. However, this is a most unfortunate misreading of the history of linguistics. Jones neither initiated the comparative method nor discovered Indo-European, as a comparison of a remarkably similar quote from Andreas Jäger in 1686, one hundred years earlier, reveals:

> An ancient language, once spoken in the distant past in the area of the Caucasus mountains and spreading by waves of migration throughout Europe and Asia, had itself ceased to be spoken and had left no linguistic monuments behind, but had as a "mother" generated a host of "daughter languages," many of which in turn had become "mothers" to further "daughters." (For a language tends to develop dialects, and these dialects in the course of time become independent, mutually unintelligible languages.) Descendants of the ancestral languages include Persian, Greek, Italic (whence Latin and in time the modern Romance tongues), the Slavonic languages, Celtic, and finally Gothic and the other Germanic tongues. (Jäger 1686, cited by Metcalf 1974: 233)

In fact, there were several notable predecessors to Jones (in addition to the supporters of the Scythian hypothesis mentioned above). For example, Edward Lhuyd (1707) compared several Indo-European languages (Celtic, Germanic, Slavic, Persian, etc.), presenting a long list of cognates, sound correspondences, and sound changes. He even discovered part of Grimm's law (which has to do with sound correspondences between Germanic and the other Indo-European languages), long before Rask and Grimm made it famous (see below). Johannis (János) Sajnovics (1770) demonstrated the relationship between Hungarian, Lapp, and Finnish. He used clear methods which were followed frequently in later work, and his work was very influential in the subsequent development of historical linguistics. For example, Rasmus Rask (1787–1832) (1993 [1818]: 283), famous early Danish historical linguist, felt confident of the evidence he presented for the kinship of Germanic with Greek and Latin because it compared favorably with Sajnovics's "proof that the Hungarian and Lappish languages are the same," which, Rask said, "no one has denied since his day." Some Africanists cite Abbé Lievin Bonaventure Proyart's *Histoire de Loango, Kakongo, et autres royaumes d'Afrique* from 1776 as a rival to Jones for its historical linguistic clarity. He pointed out that Kakongo and Laongo differ in many respects from Kikongo, but that "several similar articles [presumably prefixes], and a great number of common roots, seem, however, to indicate that these languages had a common origin" (quoted by Gregersen 1977: 97). Before Jones's famous pronouncement was published (in 1798), Jonathan Edwards, Jr. (1745–1826) (1787) demonstrated the family relationship among the Algonquian languages; Edwards listed

"some 60 vocabulary items, phrases, and grammatical features"; Jones, in contrast, presented no linguistic evidence.

Connections among Indo-European languages had been observed by many before Jones. Also, the relationship between Sanskrit and other Indo-European languages, which is generally attributed to Jones, also had already been observed by several others. For example, De Guignes (1770: 327) reported that "an infinity of travellers have already noticed that in the Indian languages and even in Sanskrit, the learned tongue of these peoples, there are many Latin and Greek words" – Jones cited de Guignes and also referred to the Scythian hypothesis.

In fact, Jones had little interest in linguistics. His plan was to write a history of peoples of Asia, and language was only one source of information, used together with information from philosophy and religion, remains of sculpture and architecture, and documents of sciences and arts (Jones 1798: 421). His interest in the history of the human "races" rather than in language was typical in eighteenth- and nineteenth-century scholarship, shared by Leibniz, Hervás y Panduro, Monboddo, Vater, Schlegel, Grimm, Humboldt, among others. Their linguistic comparisons were just part of a broader history of the nations and races of the world. This theme of language in concert with other sources of evidence to determine the history and classification of nations and races was to persist into the early twentieth century. In fact, with this orientation, Jones incorrectly classified many languages, both Indo-European and non-Indo-European ones. Nevertheless, Jones was famous before he went to India as a judge; he had written a famous Persian grammar and was renowned for his scholarship involving numerous oriental languages. People expected big things of him, and indeed through his translations of Hindu legal texts he made Sanskrit well known in Europe. As a result, his contribution came to be interpreted too enthusiastically.

Rather than being the initiator of Indo-European and of methods of comparative linguistics, Jones reflected the thinking of his day. For example, Christian Jakob Kraus (1753–1807) (1787) reviewed the assumptions concerning the comparative study of languages at that time. He indicated that similarity of words alone may or may not be indicative of family relationship, but if the grammatical structures of compared languages contained far-reaching similarities, the conclusion was in favor of a genealogical relationship (Hoenigswald 1974: 348). Very influential, and much more sophisticated than Jones' work, was Sámuel Gyarmathi's (1751–1830) *Affinitas linguae Hungaricae cum linguis Fennicae originis grammatice demonstrata* (1799), which both reflected and led the intellectual concerns of the day, emphasizing grammatical comparisons. Holgar Pedersen (1867–1953) (1962 [1931]: 105), in his famous history of linguistics, considered Gyarmathi's comparative grammar "the principle which became the lodestar of incipient Indo-European linguistics," the key to "comparative grammar." Notably, Gyarmathi warned against arguing for a genetic relationship based on similarities due to universal grammar:

> it is beyond dispute that there are universal syntactic rules shared by most nations . . . I believe that it is much more appropriate for my demonstration to bring up the kind of examples which are specifically found in Hungarian, Lapp and Finnish and which can hardly be expressed at all in Latin, German and other European languages. (Gyarmathi 1983 [1799]: 33)

With Friedrich von Schlegel [1772–1829] (1808), "comparative grammar" became a continuing focus of historical linguistic studies. Schlegel drew from biology and comparative anatomy, and employed the notion of a family tree. Grammatical structure was his main criterion of family relatedness; two languages were considered related only when their "inner structure" or "comparative grammar" presents distinct resemblances (Schlegel 1808: 6–7).

Rasmus Rask wove together the historical linguistic currents leading to his day and laid out explicitly "the principles one considers it most proper to follow" (1993 [1818]: 9). He stressed the importance of comparing grammatical structures according to Sajnovics's and Gyarmathi's methods, applying etymological principles to the genetic classification of languages (Rask 1993

[1818]: 11; Diderichsen 1974: 301). As Rask explained, "grammatical agreement is a much more certain sign of kinship or basic unity" (Rask 1993 [1818]: 33–4), but he also relied on sound correspondences and basic vocabulary as evidence (p. 34). Rask discovered the set of sound correspondences which later became known as Grimm's law (though Rask's version seems somewhat clumsy in hindsight; Rask 1993 [1818]: 161–2).

Grimm's law was a major milestone in the history of Indo-European and thus also in historical linguistics. Jakob Grimm [1785–1863], of Brothers Grimm fairytale fame, is one of the largest luminaries in historical linguistics. In the second edition of his *Deutsche Grammatik* (1822) he included the section inspired by Rask's formulation of sound correspondences among Indo-European languages later called "Grimm's law." Grimm recognized the importance of sound correspondences as evidence of family relationships, saying his law had "important consequences for the history of the language and the validity of etymology" and that it "provided sufficient evidence for the kinship of the languages involved" (Davies 1992: 161). Grimm's law treats a series of changes in certain consonants from Proto-Indo-European to Proto-Germanic:

p, t, k became f, θ [like "th" of *thing*], h, respectively
b, d, g became p, t, k, respectively
bh, dh, gh became b, d, g, respectively

(Not all the consonants involved are mentioned here.)

Some examples which illustrate Grimm's law are seen in Figure 6.1, where the words in English (a Germanic language) show the results of the changes, whereas their cognates in French (not Germanic) did not undergo the change.

While Grimm's law accounts for the systematic correspondences between Germanic and non-Germanic languages, it had some exceptions. However, subsequent discoveries, in 1862, showed that these exceptions have satisfactory explanations, and this led to a major development in linguistics. In Sanskrit and Greek, as a result of Grassmann's law, two aspirated stops within a word regularly dissimilated so that the first lost its aspiration (bh, dh, gh became b, d, g, respectively), and as a consequence, some sound correspondences between Sanskrit and the Germanic languages do not match expectations from Grimm's law, as seen in Figure 6.2.

	French	*English*
*p > f	pied	foot
*t > θ:	trois	three
*k > h:	cœur	heart
*d > t:	dent	tooth (< tanθ)
*g > k:	grain	corn
*bh > b:	frère (from *bhrā́ter)	brother

Figure 6.1 Examples illustrating Grimm's law.

Proto-Indo-European	Sanskrit	Gothic	
*bheudha-	bōdha	biudan	"to wake, become aware, bid"

Figure 6.2 Example illustrating Grassmann's law.

Proto-Indo-European	Sanskrit	Gothic	
*septém-	saptá	sibun	"seven"

Figure 6.3 Example illustrating Verner's law.

In Sanskrit, the **bh* dissimilated to *b* due to the **dh* in this word (giving Sanskrit *b* though *bh* would have been expected). In the Gothic cognate, which means "to bid", by Grimm's law we expect the *b* of the Sanskrit word to correspond to *p* in Gothic, and we expect the Gothic *b* to correspond to Sanskrit *bh*. This exception to Grimm's law is explained by the fact that Grassmann's law deaspirated the first aspirated consonant in Sanskrit. In 1877 Karl Verner (1846–96) accounted for other exceptions to Grimm's law in a change known as Verner's law, illustrated in Figure 6.3.

By Grimm's law, we expect the *p* of Sanskrit to correspond to *f* in Gothic, not the *b* found in this Gothic word, and given the *b* of Gothic, we would expect Sanskrit to have *bh*. Verner's law explains this exception to Grimm's law. When the Proto-Indo-European accent followed the sound in question (and it was not the first sound in the word), as seen in Sanskrit *saptá* (*á* is accented), **p* became *b* in Germanic, as in the Gothic word; otherwise, Grimm's law applied.

4.3 The Neogrammarians

This success in accounting for what had originally appeared to be exceptions to Grimm's law spawned one of the most notable developments in linguistics. It led the Neogrammarians to the confidence that sound change was regular and exceptionless. The Neogrammarians, beginning in about 1876 in Germany, became extremely influential. They were a group of younger scholars who antagonized the leaders of the field by attacking older thinking and loudly proclaiming their own views. They were called *Junggrammatiker* "young grammarians" in German, where *jung-* "young" had the sense of "young Turk," originally intended as a humorous nickname for these rebellious and strident young scholars, although they adopted the name as their own. They included Karl Brugmann (1849–1919) (the most famous linguist of his time), Berthold Delbrück (1842–1922), August Leskien, Hermann Osthoff (1847–1909), Hermann Paul (1846–1921), and others. The Neogrammarian slogan, "sound laws suffer no exceptions," or, more precisely, "every sound change, in as much as it occurs mechanically, takes place according to laws that admit no exceptions," was declared virtually as doctrine in the so-called "Neogrammarian manifesto" of Hermann Osthoff and Karl Brugmann (1878), written mostly by Brugmann. This became an important cornerstone of reconstruction by the comparative method. By "sound laws" they meant merely "sound changes," but referred to them as "laws" because they linked linguistics with the rigorous sciences which dealt in laws and law-like statements.

Some scholars, many of them dialectologists, did not accept the Neogrammarian position that sound change is regular and exceptionless, but rather opposed this and the "family tree model" which represents languages related by virtue of descent from a common ancestor. The "family tree model" is often associated with August Schleicher (1821–1868), prominent pre-Neogrammarian figure in Indo-European linguistics (see Schleicher 1861–2). This model is typically linked in the literature with the development of the comparative method and eventually with the Neogrammarian notion of the regularity of sound change (though this connection is not necessary). The opponents' slogan was "each word has its own history." This slogan is often attributed to Jules Gilliéron (1854–1926), author of the *Atlas linguistique de la France* (1902–10), the dialect atlas of France, although it really comes from Hugo Schuchardt (1842–1927), a contemporary of the early Neogrammarians, of whose claims he was critical. The alternative to the family tree model which was put forward was the "wave theory," usually attributed to Johannes Schmidt (1872) though it, too, was actually developed earlier, in 1868 and 1870, by Hugo Schuchardt (Alvar 1967:

82–5). Interestingly, Schuchardt and Schmidt were both students of Schleicher, as were several of the leading Neogrammarians. The "wave theory" was intended to deal with changes due to contact among languages and dialects, where changes were said to emanate from a center as waves on a pond do when a stone is thrown into it, and waves from one center of dispersion (where the stone started the waves) can cross or intersect outward moving waves coming from other dispersion centers (started by other stones thrown into the water in other locations). Changes due to language contact (borrowing) were seen as analogous to successive waves crossing one another in different patterns. The dialectologists' slogan, that every word has its own history, reflects this thinking – a word's history might be the result of various influences from various directions, and these might be quite different from those involved in another word's history; hence each word has its own (potentially quite different) history.

Although some scholars have thought that dialectology naturally led to challenges to the Neogrammarian position, in fact the Neogrammarian founders gained support for their position in dialect study. They were impressed by Jost Winteler's (1876) study of the Kerenzen dialect of Swiss German in which he presented phonological statements as processes, modeled after Pāṇini's ancient rules for Sanskrit. This "regularity" which Winteler saw in the dialect's modern rules – for example that in Kerenzen every *n* became ŋ [like "ng" in English *sing*] before *k* and *g* – inspired them to have confidence in the exceptionlessness of sound changes (Weinreich, Labov, and Herzog 1968: 115). Today it is recognized that both the family tree and the wave model are necessary to explain change and that they complement one another (Campbell 1998: 187–91).

5 Philosophical-Psychological (-Typological-Evolutionary) Approaches

While the Neogrammarian tradition has dominated the history of linguistics, there was another once influential orientation, a philosophical-psychological-typological-evolutionary outlook on the nature and evolution of language, now largely forgotten.

In the nineteenth century, there was a clash between views of linguistics as a "*Naturwissenschaft*" (physical science) and "*Geisteswissenschaft*" (humanities). Leading linguists attempted to place linguistics in the natural (physical) sciences, denying any value for the more humanistic, "sentimental" intellectual orientations. A close analogy of linguistics with biology had been insisted upon by Schlegel, Rask, and many others, a view associated especially with Schleicher (1861–2). Nevertheless, many in the past did not clearly separate language, race, nation, and culture. As seen above, Jones, Leibniz, Hervás y Panduro, Adelung, Rask, and others believed they were working out the history of races and nations in their linguistic works, rather than that of mere languages. Folk (or national) psychology, coupled with the assumed stage of social evolution attained by its speakers – often called "progress" – was thought to determine a language's typology and its history, the sort of gross linguistic history later eschewed by the mainstream as too psychological. In the eighteenth century, interest began to concentrate on the origin of differences in languages and cultures, and this led to the idea of the particular "genius" of each language and through this to a "typology" of languages. These types were often viewed as both descriptive and historical. Traditional etymology and theories of language relationship were merged with logical grammar in an evolutionary scheme. Languages were classified into types according to their morphological structure, the types taken as representing or being correlated with evolutionary stages. Structural change in a language was taken as nothing more than the result of social evolution. For many, following Wilhelm von Humboldt (1767–1835), the typological categories – isolating, agglutinative, flexional, and incorporating – were taken as reflecting the level of social evolution attained by the speakers of the language (a typical equation was: isolating = savagery, agglutinative = barbarianism, inflectional = civilization). For example, for

Friedrich Müller (1834–1898), social evolution, racial type, and language type were correlated, so that hair shape and linguistic morphology (structure of words) could be associated with one another.

The notion of "inner structure" was persistent in this orientation. Johann Gottfried von Herder (1744–1803) (1772) had spoken of the "inner development of language," and the notion of "inner structure" was prominent in the work of Adelung, Schlegel, Bopp, Humboldt, Steinthal, and others.

Franz Bopp's (1791–1867) (1816, 1833–52) comparative grammar contributed significantly to growing interest in comparative grammar, but also incorporated aspects of the philosophical-psychological-typological-evolutionary outlook. Schleicher's (1861–2) *Compendium der vergleichenden Grammatik der indogermanischen Sprachen* is the acknowledged synthesis of nineteenth-century comparative linguistics for its time. Schleicher followed Humboldt's (1822) types, expounding the view that languages evolve, or "progress," from isolation to agglutination (with affixes arising from full words) and move onward to flexion, with gradual progress in the direction from simple to complex forms. Schleicher believed that "growth" (through agglutination) took place only in the prehistoric phase when languages were still young and capable of word formation, during the period of *Sprachbildung* ("language formation"), whereas only changes of "decay" (by sound change and analogy) took place in the later historical period, after the growth process was assumed to have ceased entirely, during the period of *Sprachgeschichte* ("language history").

This view, that modern languages are but dim reflections of their more perfect progenitors, was called "glottogonic"; it characterizes the work of many early comparativists, but was severely criticized by Neogrammarians. They rejected Schleicher's and others' orientation as "glottogonic speculation." They denied its separation of stages, insisting that the same kinds of language changes apply to all phases of linguistic history; analogy and sound change operate throughout a language's history (Paul 1920 [1880]: 174; see Davies 1986: 154; Harris and Campbell 1995: 17–19).

Aspects of the philosophical-psychological-typological-evolutionary outlook endured into the twentieth century, although it was played down in the official histories written mostly by Neogrammarians, e.g. Pedersen (1962 [1931]); see Boas, Sapir, and Whorf, below; Campbell (1997: 27–9, 37–43, 55–66).

6 The Rise of Structuralism

Thinking which led to the replacement of the historical orientation in linguistics by emphasis on the study of living languages and their structure came from a number of quarters at roughly the same time. For example, incipient notions of the "phoneme" developed in several areas at about the same time, so that it is not possible to attribute it to any one person or school. The phoneme is a central concept of linguistics whose definition varies from school to school but which basically means the significant units of sound, the minimal unit of sound capable of changing the meaning of words. Some speculate that, in the wake of World War I, linguists were happy to free themselves of the German domination represented by the Neogrammarian historicism which had been predominant until then, and indeed the new currents, partly convergent but also with individual characteristic differences, came not from Germany, but from Switzerland with de Saussure, Russia with Baudouin de Courtenay, and America with Boas.

6.1 Ferdinand de Saussure (1857–1913)

After early influential Neogrammarian work on the vowels of Indo-European in 1878, published when he was 21, and a doctoral dissertation in 1881 on the genitive in Sanskrit, Saussure published little else, nothing on the topics for which he is best known, and yet he became one of

the most influential scholars in twentieth-century linguistics and modern intellectual history. The extremely influential *Cours de linguistique générale* (1959 [1916]), published after his death in 1913, was compiled from his students' notes from his course in general linguistics (given three times between 1907 and 1911) at the University of Geneva. This book is credited with turning the tide of linguistic thought from the diachronic (historical) orientation which had dominated nineteenth-century linguistics to interest in the synchronic (nonhistorical) study of language. Defining linguistics was a main goal of the book.

Saussure emphasized the synchronic study of language structure and how linguistic elements are organized into the system of each language. His theory of signs has been very influential. His linguistic sign is a union of the *signifiant* ("signifier," the form, sound) and the *signifié* ("signified," the meaning, function); the particular form (sounds) and the particular meaning in individual signs are arbitrarily associated with one another; their connection is purely conventional; that is, the sound–meaning association in signs is not predictable from one language to the next. The thing signified, say the notion *tree*, is arbitrarily associated with the sounds (signifier) which signal it, for example with the sounds of *Baum* in German, *kwawitl* in Nahuatl, *rakau* in Maori, *tree* in English, and so on. In Saussure's view, linguistic entities were considered members of a system and were defined by their relations to one another within that system. He compared language to a game of chess, a highly organized "algebraic" system of relations, where it is not the actual physical attributes of the pieces which define the game, but rather the relation of each piece to the other pieces in the system which give it its definition, a system *où tout se tient* ("where everything holds together," where everything depends on everything else, that is, where everything is defined in terms of its relations to everything else), in the famous saying of Antoine Meillet (1866–1917) (student of Saussure).

Saussure, influenced by the social thinking of Émile Durkheim (1858–1917) (a founding figure in sociology), held that language is primarily a "social fact" (rather than a mental or psychological one, as others had held), that is, that there is a "collective consciousness" which is both the possession of society at large but also defines society. ("Social fact" and "collective consciousness" are terms associated with Durkheim, which Saussure used.) Saussure's famous dichotomy, *langue* (language, as socially shared and as a system) versus *parole* (speech, the language of the individual), reflects the French social thinking of the day. The goal, naturally, was to describe *langue*, but, since the individual's speech would reflect and represent the language as possessed by society generally, the social (general) character of language could be approached through the study of the language of the individual.

Today, nearly all approaches to linguistics are "structuralist" in some sense and reflect Saussure's monumental influence. Saussure's structuralism has also had a strong impact on anthropology, literary criticism, history, psychology, and philosophy, promoted and modified by Jakobson, Lévi-Strauss, Foucault, Barthes, and Derrida, among others.

6.2 *The Prague School and its antecedents*

Jan [Ignacy Niecisław] Baudouin de Courtenay (1845–1929), born in Poland, was developing structuralist ideas at the University of Kazan in Russia at about the same time as Saussure was lecturing in Geneva. Saussure was familiar with Baudouin de Courtenay's thinking and parts of the *Cours* reflect this very directly; Saussure had said that Baudouin and his student Mikołaj Kruszewski (1851–1887) were the only European scholars who contributed to linguistic theory (Stankiewicz 1972: 4–5). Baudouin de Courtenay's thinking was instrumental in the development of the notion of the "phoneme," though the concept developed with influence also from several other directions at once. Baudouin and his students contributed the terms "morpheme," "grapheme," "distinctive feature," and "alternation," all basic terminology in modern linguistics. His thinking survived most vividly through linguists whom he influenced who became associated with the Linguistic Circle of Prague.

Serge Karcevskij (1884–1955), who had been in Geneva from 1906 to 1917, brought Saussure's thinking back to the Moscow Linguistic Circle, with its formalist movement. Roman Jakobson (1896–1982) and Prince Nicholai S. Trubetzkoy (1890–1938) recognized areas of convergent thinking with Saussure. Later, Jakobson and Trubetzkoy (two Russians) became the best known representatives of the Prague School of linguistics. Jakobson, Trubetzkoy, and others of the Prague School developed aspects in structuralism which are important in current theories, for example "distinctive features," "markedness," "topic," and "comment," and the notion of "implicational universals," as well as "linguistic areas" (*Sprachbund*). Jakobson, who emigrated to the US in 1942, had a strong impact on the development of generative phonology both through his student, Morris Halle, and through his influence on Noam Chomsky (see below).

6.3 Franz Boas (1858–1942)

Franz Boas is considered the founder of American linguistics and American anthropology. A major concern for him was to obtain information on Native American languages and cultures before they disappeared, and indeed his is the last, sometimes the only, significant data on a number of now extinct languages, for example, Lower Chinook, Cathlamet, Chemakum, Pentlach, Pochutec, and Tsetsaut. He passed his sense of urgency for fieldwork on to his students, a dedication to getting accurate information while it was still possible. The methods Boas and his followers worked out for the description of such languages became the basis of American structuralism, a dominant force in twentieth-century linguistics.

This approach reflects Boas's famous "linguistic relativity" and his emphasis on avoiding generalization. At that time, many erroneous claims were about, such as that certain South American Indians could not communicate in the dark, since, it was asserted, their language was so "primitive" they had to rely on gestures (which could not be seen in the dark) to convey concepts such as "here" and "there" or "yesterday" and "tomorrow" to make up for the assumed severe limitations of their vocabulary; that change in "primitive" languages could proceed so fast that grandparents could not understand their grandchildren; that the pronunciation of "primitive" languages could vary unpredictably and be so imprecise as to make learning such languages all but impossible; and so on. In particular, earlier descriptions of so-called "exotic" languages frequently attempted to force them into traditional European grammatical categories, missing or distorting many distinctions significant to these languages. The different categories available in human languages are far more extensive than had been supposed from the generalizations being made which were based on the more familiar European languages. In face of so many bad generalizations, Boas believed it important to avoid preconceptions and to describe each language and culture in its own terms – on the basis of information derived internally from an analysis of the language itself rather than imposed on it from outside. His students made this a matter of principle, an orientation to linguistics with emphasis on description and against generalizing, against theorizing about language. This orientation prevailed in American Structuralism until Noam Chomsky's views reoriented the field towards universals, generalizing, and linguistic theory (see below).

The notion of "inner form" became the core of Boas's view of ethnology and linguistics. Boas used Humboldt's concept of "inner form" to deal with the diversity of American Indian languages, seeing languages as conditioning the world view of their speakers. He was strongly opposed to the evolutionism of philosophical-psychological-typological-evolutionary views of the past, but maintained a Humboldtian psychological orientation. Nevertheless, he succeeded in turning attention against the evolutionary determinism characteristic of this way of thinking. He showed that the traditional typological-evolutionary views of grammar were inaccurate and ethnocentric. His view is revealed in his conception of the *Handbook of North American Indian Languages* (Boas 1911) as a "morphological classification" of American Indian languages. The languages he selected for inclusion in the *Handbook* were chosen to represent as many

psychologically distinct types of language as possible, with the goal to reveal their "morphological classification and psychological characterization" and to serve as "a uniform series of outlines of Indian languages to be published in synoptic form for use in comparative studies by the philologists [historical linguists] of the world." "His emphasis was on the diversity of linguistic structures and accompanying mental worlds to be found in North America" (Campbell 1997: 64). After Boas, with help from Sapir and Kroeber, the view of morphological types as representatives of stages of social evolution died out. The two most influential American linguists after Boas were Sapir and Bloomfield.

6.4 Edward Sapir (1884–1939)

Sapir (Boas's student) was highly admired during his life and is still something of a hero to many linguists. He published extensively in both linguistics and anthropology, did firsthand fieldwork on many American Indian languages, contributed to historical linguistics (in Indo-European, Semitic, and numerous Native American families; for example, he established once and for all the Uto-Aztecan family and proposed the once controversial but now established Ritwan-Algonquian family), and wrote theoretical works, for example on the phoneme, still read with profit today. His impact in these areas was monumental. At the same time, he was also no stranger to the psychological-typological current of thought. Trained in Germanic linguistics, he fully understood the Humboldtian psychological tradition. His 1921 book, *Language*, insightfully dealt with the broad morphological typologies of the past century, but without the evolutionism which characterized them in earlier views. His own typology rested on the tradition extending from the eighteenth and nineteenth centuries represented by Schlegel, Bopp, Humboldt, Schleicher, Müller, Steinthal, Wundt, and others. However, like Boas, he rejected the evolutionary prejudice that typified traditional typological studies: "all attempts to connect particular types of linguistic morphology with certain correlated stages of cultural development . . . are rubbish" (Sapir 1921: 219). He did not accept the notion of significant racial differences in the "fundamental conformation of thought," the belief that differences in linguistic forms (believed to be connected with the actual processes of thought) could be indexed to racial differences. However, he did uphold the psychological orientation of the earlier typological tradition and passed it along to his student Benjamin Whorf (1897–1941), in whose hands it was transformed into the Whorf (or Sapir–Whorf) hypothesis, which holds that a speaker's perception of the world is organized or constrained by the linguistic categories his or her language offers, that language structure determines thought, how one experiences and hence how one views the world. This became a lasting theme in linguistics, anthropology, psychology, and philosophy, though many are unaware of its pedigree from German Romanticism. In his descriptive work, Sapir maintained the mentalism and nongeneralizing of Boas's approach.

6.5 Leonard Bloomfield (1887–1949)

Bloomfield is credited with giving American structuralism its fundamental form, making linguistics an autonomous field. His principal concern was to develop linguistics as a science. Bloomfield's (1933) *Language* is considered a milestone in linguistics, the foundation of American structuralist linguistic thinking. Of this book, Bloomfield reported that it showed Saussure's thinking on every page. Bloomfield was also heavily influenced by behaviorist psychology. He accepted the Boasian prohibition against generalizing but at the same time he denied the relevance of "mind"; that is, he opposed the mentalism that had characterized the American linguistics of Boas, Sapir, and their students. This left American structuralism (represented by Bernard Bloch, Zellig Harris, Charles Hockett, Henry Lee Smith, George Trager, and others – sometimes called the "Bloomfieldians") with essentially nothing more than method, the "discovery procedures" against which Chomsky later argued so effectively. With a mentalistic orientation but

no theoretical assumptions (no generalization), followers of Boas and Sapir could hold their description of a given language up to some external measure to decide whether it was accurate or not, namely, by determining whether it reflected what native speakers knew of their language. However, Bloomfield and his followers were left with no means of validating a description – by denying generalizations (theory), they could not evaluate the description of a given language according to how well it conformed to an understanding of human language in general, and by denying "mind" (mentalism) they could not judge a description against the extent to which it matched what native speakers knew of the structure of their language. Thus, nothing remained except method, "discovery procedures," the search for contrast and complementary distribution in the data recorded by linguists. This is a particularly impoverished state for a "science" to find itself in – all method and no theory. Given this state of affairs, it is not surprising that Chomsky was able to bring about a revolution in linguistics.

7 Noam Chomsky and Linguistic Theory since 1957

The mainstream of linguistics since 1957, the year in which Chomsky's *Syntactic Structures* appeared, has been dominated by Noam Chomsky (1928–). It is difficult to overestimate Chomsky's impact on both linguistics and contemporary ideas in general: "Chomsky is currently among the ten most-cited writers in all of the humanities [and social sciences] (behind only Marx, Lenin, Shakespeare, the Bible, Aristotle, Plato, and Freud) and the only living member of the top ten" (Pinker 1994: 23). It is common to speak of "the Chomskian revolution," so radically distinct is Chomsky's program from that of his American structuralist predecessors. Unlike the Bloomfieldians, Chomsky brought back mentalism. For him, the goal of a grammar is to account for the native speaker's "competence," defined as what a native speaker knows, tacitly, of his or her language. Since speakers know, among other things, how to produce an infinite number of sentences, many of which are novel, never having been produced before (talked about as linguistic "creativity"), an account of "competence" would require the formal means to produce or generate these new sentences, hence a "generative grammar." A grammar was seen as a theory of a language, constrained and evaluated just as any other theory in the sciences. Unlike most of his predecessors, Chomsky focussed on syntax, and in so doing, laid the foundation for explaining this "creativity." The notation of generative grammar was invented to make explicit the notion of "competence"; a generative grammar is a formal system (of rules, later of principles and parameters) which makes explicit the finite mechanisms available to the brain to produce infinite sentences in ways that have empirical consequences and can be tested as in the natural sciences.

Unlike the Boasians and the Bloomfieldians, Chomsky gave linguistics the goal of generalizing, of attempting to determine what languages hold in common and to establish a rich theory of human language. Chomsky's approach is often called "generative grammar" or "transformational-generative grammar (see Chapter 7)." Transformations were essentially rules for relating one syntactic structure to another, for example, as in early versions where questions, such as *Is Pat here?*, were derived by transformation from the corresponding declarative, *Pat is here*. However, in later versions of the theory, transformations no longer play a significant role. In Chomsky's theorizing about language, universals hold a central place. He rejected the "discovery procedures" of his American structuralist predecessors, those inductive procedures for deriving the grammatical description of a language through the application of procedures sensitive essentially only to the distribution of elements in a corpus of data from the language. The primary task of the linguist, according to Chomsky, should not be to discover the structure of the language from a body of data; rather, the goals should be to describe and explain the knowledge of the structure of the language which the native speaker has. This shifted attention from actual behavior (or recorded data) to the system of knowledge that underlies the production and understanding of language, and, further, to the general theory of human language lying behind this knowledge.

This was a radical reorientation of the field, rejecting the antimentalism of the Bloomfieldians and the antitheorizing of the Boasians and Bloomfieldians.

Chomsky redirected the goal of linguistic theory towards attempting to provide a rigorous and formal characterization of the notion "possible human language," called "Universal Grammar." In his view, the aim of linguistics is to go beyond the study of individual languages to determine what the universal properties of human language in general are, and to establish the "universal grammar" that accounts for the range of differences among human languages. The theory of grammar relies on certain general principles which govern the form of the grammar and the nature of the categories with which it operates. These principles are conceived of as universal properties of language, properties that are biologically innate. The notion of innateness, developed by E. H. Lenneberg (1960), was adopted by Chomsky and became central to his thinking. He argued that much of our knowledge about language is universal and innate, that is, inborn, genetically endowed, a language instinct, part of our biological birthright. Chomsky attacked a standard view at the time that children are born with minds that are essentially "blank slates" (the view of the behaviorist psychologists), that the human psyche is largely molded by the surrounding culture. Chomsky maintained that, rather than being born blank slates, children have a genetic predisposition to acquire linguistic knowledge in a highly specific way. He posited innate principles that determine the form of acquired knowledge.

Chomsky's (1959) review of Skinner's *Verbal Behavior* (1957) became the basic refutation of behaviorist psychology (which had so influenced Bloomfield and his followers). B. F. Skinner (1904–1990) had claimed to be able to explain language as a set of habits gradually built up over the years – as in experiments with rats rewarded with pellets of food in their trial-and-error learning (*operant conditioning*), which Skinner assumed to be the mechanism by which the vast majority of human learning takes place, including language learning. Understand the "controlling variables" (stimuli) and responses, and you understood language learning, he claimed. Chomsky's criticism showed that rat behavior is irrelevant to human language learning and that Skinner had misunderstood the nature of language. Human utterances are not predictable in face of a particular stimulus; we might not say only "oh what a beautiful picture" when seeing a painting, but also, "it clashes with the wallpaper," "it's hanging too low," "it's hideous," etc. In caretaker–child interactions, says Chomsky, parents approve / reward statements which are true rather than those which are grammatically correct. A child's ungrammatical utterance, "Teddy sock on," is approved by the mother when the child shows her a teddy bear wearing a sock, but "Look, Teddy is wearing a sock" receives the mother's disapproval when the child shows the mother a bear without a sock. Perhaps some human activities, say learning to drive or to knit, may seem to be learned as the rats learned, but not language. Language structure is very complex, but children do not go through a prolonged trial-and-error phase. In Chomsky's words:

> A consideration of the character of the grammar that is acquired, the degenerate quality and narrowly limited extent of the available data, the striking uniformity of the resulting grammars, and their independence of intelligence, motivation, and emotional state, over wide ranges of variation, leave little hope that much of the structure of the language can be learned by an organism initially uninformed as to its general character. (Chomsky 1964: 58)

As evidence of innateness, the following have been offered. Language is extremely complex but children acquire it in a remarkably short period of time. The stimulus or experience children have with the language around them appears to be too poor to provide the basis for acquiring the mature linguistic capacities that they ultimately attain. The language around them that children experience consists partly of degenerate data which have little effect on the capacity which emerges; the speech children hear is full of unfinished sentences, mistakes, slips of the tongue (performance errors). It contains few or no example sentences to illustrate some of the complex structures that children "learn." Children's experience is finite (limited), but the capacity

eventually attained includes the ability to produce an infinite number of grammatical sentences. This is often called "the poverty of stimulus argument." The acquisition of language is relatively independent of intelligence – the language learning ability of the least able children is not noticeably inferior to that of bright children; all but those with the most severe disabilities learn language, and language emerges at about the same time in children all over the world, uniformly regardless of language environment, culture, or ethnicity. Skill or ability seem to have nothing to do with it; however, for most other learned tasks, like roller-skating, piano-playing, etc., there are enormous differences from child to child. Finally, the direct correction of children's language mistakes (as Skinner's model advocates) has been noted by numerous researchers to be pointless; children's production changes not with adult correction, but only as the grammar they acquire goes through the normal stages of language development in children.

Since this theory began, it has evolved through versions called "Standard Theory," "Extended Standard Theory" (and "The Lexicalist Hypothesis"), "Trace Theory," "Government and Binding" (later called "Principles and Parameters" approach), and finally "the Minimalist Program." It has also spawned a number of theories which compete in some ways but which nevertheless share most of the Chomskyan goals of linguistics and many of the underlying assumptions, for example, "Case Grammar," "Generalized Phrase Structure Grammar," "Generative Semantics," "Head-Driven Phrase Structure Grammar," "Lexical-Functional Grammar," and "Relational Grammar."

8 Typology

An orientation to linguistics which contrasts with the "generativist" approach is that of the "typologists," sometimes called the "functional-typological" or "Greenbergian" approach. Typology, broadly speaking, is the classification of languages according to linguistic traits and the comparison of patterns (structures) across languages. The typological approach attempts to explain the patterns through appeal to language function in cross-linguistic comparison. Languages can be typologized according to almost any linguistic trait, and indeed classifications based on widely varied attributes have been proposed in the history of linguistics. For example, Wilhelm Wundt (1832–1920) (1990: 436) dealt with 12 oppositions or types, including prefixing versus suffixing languages, free versus fixed word-order languages, and languages with more extensive grammatical apparatus of verbs versus those with more elaborate treatment for nouns. Such typologies rest on a tradition extending from the eighteenth and nineteenth centuries represented by Schlegel, Bopp, Humboldt, Schleicher, and others. Typology throughout the nineteenth century was primarily morphological – the structure of the word (morphology) alone was held to determine a language's whole character.

Several concepts fundamental to modern approaches to typology come from the Prague School, for example, implicational universals – if a language has a trait *x*, then it is expected also to have a trait *y*; for example, the presence of nasalized vowels in a language implies that language will also have plain, non-nasalized vowels. Roman Jakobson (1958) brought implicational universals to broader attention and this marks the beginning of modern work on typology and universals. It inspired Joseph H. Greenberg's (1915–) classic article on word order (1963); Greenberg is generally considered the founder of modern typology. Typological study has contributed to the understanding of many concepts of grammar and of how they interact with one another, how they function, and how they are distributed in the world's languages. Typological research also incorporates many assumptions about how languages can change, and "grammaticalization" has become the subject of extensive discussion. Though notions of grammaticalization have a long earlier history in linguistics (Harris and Campbell 1995: 15–20, 45–7), Antoine Meillet (1912) introduced the term, which has come to mean primarily changes in which an independent

word changes to become a grammatical marker, or where some less grammatical entity becomes more grammatical. A standard example is change from *will* with its original meaning of "want" to the grammatical "future tense" (Traugott and Heine 1991: 2).

9 Conclusions

In a brief survey of the history of linguistics such as this, much of significance goes unmentioned, though the major developments have been described here. Suffice it to say that linguistics is commonly held to be one of the most successful of the social sciences and as such has contributed both methods and models of rigor to other disciplines. As well as having its own robust history, linguistics has contributed richly to the general history of ideas and can be expected to continue to do so. Therefore, to conclude, it may be appropriate to attempt to anticipate the future, what the continuing history of linguistics will bring. We may guess from the current "hottest" topics in linguistics what some future areas of high activity may be. Endangered languages will continue to be a major concern – languages are becoming extinct at an alarming rate; it is estimated that within the next hundred years, 50 percent to 85 percent of the world's 6,000 or so languages will become extinct or so near to extinction they cannot be revived. Human cognition and connections with formal grammar are a major focus of the discipline, and this is likely to grow rather than diminish. Interfaces between linguistics and computer science are growing and are likely to be of high interest to future linguists. Investigation into language universals and typology, within both formal and functionalist approaches, will no doubt persist, aimed at understanding language universals, the properties of universal grammar, and the function of language (and how function may help shape language structure). The extent to which these approaches will converge or diverge even further is anyone's guess. Reports in the nonlinguistic media make the issue of remote language relationships appear to be one of the biggest concerns of present-day linguists. In fact, it is the concern of very few linguists; nevertheless, efforts to work out the history of human languages and their more distant family relationships will continue, though it is hoped that a more rigorous and careful methodology will be applied and that some progress will be made. Advances will be made in the explanation of how and why languages change. A favorite pastime of some linguists today is to speculate about what will happen to linguistics when Noam Chomsky retires and his personal influence no longer determines much of the central activity in linguistic theory. Here, speculations run rampantly in many directions. It will be fascinating to see what the future will bring.

REFERENCES

Abailard, Pierre (Peter Abelard). (1970 [ca. 1130]). In L. M. de Rijk (ed.), *Dialectica*, 2nd edn. Philosophical Texts and Studies. Assen: Van Gorcum.

Alvar, M. (1967). *Linguistica Romanica* [originally by I. Iordan, reworked and heavily annotated by M. Alvar]. Madrid: Ediciones Alcala.

Arnauld, A. and Lancelot, C. (1660). *Grammaire générale et raisonnée de Port Royal*. Paris: Pierre le Petit. (English translation 1975: *General and Rational Grammar: The Port-Royal Grammar*, edited and translated by J. Rieux and B. E. Rolling. Janua Linguarum, Series Minor, 208. The Hague: Mouton.)

Bibliander (Buchmann), T. (1548). *De ratione communi omnium, linguarum et literarum commentarius*. Zurich: C. Froschauer.

Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart and Winston.

Boas, F. (1911). Introduction. In F. Boas (ed.), *Handbook of American Indian Languages*, Vol. 1. Washington: Bureau of American Ethnology.

Bopp, F. (1816). *Ueber das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persichen und germanischen Sprache, nebst Episoden des Ramajan und Mahabharat*

in genauen, metrischen bersetzungen aus dem Originaltexte und einigen Abschnitten aus den Veda's. Frankfurt am Main: Andreische Buchhandlung.

Bopp, F. (1833–52). *Vergleichende Grammatik des Sanskrit, Zend, Armenischen, Griechischen, Lateinischen, Litauischen, Altslavischen, Cothischen und Deutschen*, 6 vols. Berlin: Ferdinand Dummler.

Borst, A. (1959). *Der Turmbau von Babel*. Stuttgart: Hiersemann.

Breva-Claramonte, M. (1983). *Sanctius' Theory of Language: A Contribution to the History of Renaissance Linguistics*. Studies in the History of Linguistics, 27. Amsterdam: John Benjamins.

Bursill-Hall, G. L. (1995). Linguistics in the later middle ages. In E. F. K. Koerner and R. E. Asher (eds), *Concise History of the Language Sciences: From the Sumerians to the Cognitivists* (pp. 130–7). Oxford: Pergamon.

Campbell, L. (1997). *American Indian Linguistics: The Linguistic History of Native America*. New York: Oxford University Press.

Campbell, L. (1998). *Historical Linguistics*. Edinburgh: Edinburgh University Press.

Chomsky, N. (1959). Review of *Verbal Behavior*, by B. F. Skinner. *Language*, 35, 25–37.

Chomsky, N. (1964). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Davies, A. M. (1986). Karl Brugmann and late nineteenth-century linguistics. In T. Bynon and F. R. Palmer (eds), *Studies in the History of Western Linguistics, in Honour of R. H. Robins* (pp. 150–71). Cambridge: Cambridge University Press.

Davies, A. M. (1992). Comparative-historical linguistics. In W. Bright (ed.), *International Encyclopedia of Linguistics*, Vol. 2 (pp. 159–63). Oxford: Oxford University Press.

De Guignes, J. (1770). *Histoire de l'Academie des Inscriptions*. Paris.

Diderichsen, P. (1974). The foundation of comparative linguistics: revolution or continuation? In D. Hymes (ed.), *Studies in the History of Linguistics: Traditions and Paradigms* (pp. 277–306). Bloomington, IN: Indiana University Press.

Droixhe, D. (1984). Avant-propos. In D. Droixhe (ed.), *Genèse du comparatisme indoeuropéen = Histoire, Épistémologie, Langage*, 6, 2, 5–16.

Edwards, J., Jr. (1787). *Observations on the Language of the Muhhekaneew Indians*. (Communicated to the Connecticut Society of Arts and Sciences, and published at the request of the Society.) New Haven: Josiah Meigs. (Reprinted London: W. Justins, Shoemaker-Row, Blackfriars; 1788, also reprinted with notes by John Pickering, in the Massachusetts Historical Society Collection, second series, 10:81–160, and Boston: Phelps and Farnham, 1823.)

Fellman, J. (1975). On Sir William Jones and the Scythian language. *Language Science*, 34, 37–8.

Gilliéron, J. and Edmont, E. (1902–13). *Atlas linguistique de la France*, 13 vols. Paris: Champion.

Gragg, G. B. (1995). Babylonian grammatical texts. In E. F. K. Koerner and R. E. Asher (eds), *Concise History of the Language Sciences: From the Sumerians to the Cognitivists* (pp. 19–21). Oxford: Pergamon.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (ed.), *Universals of Language* (pp. 73–113). Cambridge, MA: MIT Press.

Gregersen, E. A. (1977). *Language in Africa: An Introductory Survey*. New York: Gordon and Breach.

Grimm, J. (1822). *Deutsche Grammatik*. Erster Theil. Gottingen: Dieterich. (First edn published 1819.)

Gyarmathi, S. (1981 [1799]). *Grammatical Proof of the Affinity of the Hungarian Language with Languages of Fennic Origin*, translated, annotated, and introduced by V. E. Hanzeli. Amsterdam: John Benjamins. (Originally published as *Affinitas linguae Hungaricae cum linguis Fennicae originis grammatice demonstrata*. Gottingen: Johann Christian Dieterich.)

Harris, A. C. and Campbell, L. (1995). *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.

Harris, J. (1751). *Hermes, or a Philosophical Inquiry concerning Languages and Universal Grammar*. London: J. Nourse.

Herder, J. G. (1772). *Abhandlung uber den Ursprung der Sprache*. Berlin: C. F. Voss.

Hoenigswald, H. M. (1974). Fallacies in the history of linguistics: notes on the appraisal of the nineteenth century. In D. Hymes (ed.), *Studies in the History of Linguistics: Traditions and Paradigms* (pp. 346–58). Bloomington, IN: Indiana University Press.

Hoenigswald, H. M. (1990). Descent, perfection and the comparative method since Leibniz. In T. de Mauro and L. Formigari (eds), *Leffiniz, Humboldt, and the Origins of Comparativism* (pp. 119–32). Amsterdam: John Benjamins.

Hovdhaugen, E. (1982). *Foundations of Western Linguistics: From the Beginning to the End of the First Millennium AD*. Oslo: Universitetsforlaget.

Humboldt, W. von (1822). *Ueber das Entstehen der grammatischen Formen, und ihren Einfluss auf die Ideenentwicklung*. Abhandlungen der historisch-philologischen Klasse der Königlichen Akademie der Wissenschaften zu Berlin.

Jäger, A. (1686). *De Lingua Vetustissima Europae, Scytho-Celtica et Gothica*. Wittenberg.

Jakobson, R. (1958). Typological studies and their contribution to historical comparative linguistics. In E. Sivertsen, C. J. Borgstom, A. Gallis, and

A. Sommerfelt (eds), *Proceedings of the Eighth International Congress of Linguists* (pp. 17–25). Oslo: Oslo University Press.

Jones, Sir W. (1798). Third anniversary discourse: on the Hindus (February 2, 1786). *Asiatick Researches*, 1, 415–31.

Kraus, C. J. (1787). Rezension des Allgemeinen vergleichenden Worterbuches von Pallas. *Allgemeinen Literatur-Zeitung*, nos 235–7. [English translation, 1985: *Historiographia Linguistica*, 11, 229–60.]

Lenneberg, E. H. (1960). Language, evolution, and purposive behavior. In S. Diamond (ed.), *Culture in History: Essays in Honor of Paul Radin* (pp. 869–93). New York: Columbia University Press.

Lhuyd, E. (1707). *Archaeologia Britannica*. *Archaeologia Britannica, giving some account additional to what has been hitherto publish'd, of the languages, histories and customs of the original inhabitants of Great Britain: from collections and observations in travels through Wales, Cornwal, Bas-Bretagne, Ireland and Scotland*. Oxford: Printed at the Theater for the Author. (Reprinted 1969 (*English Linguistics 1500–1800*, no. 136, a collection of facsimile reprints selected and edited by R. C. Alston) Menston, England: Scolar Press.)

Locke, J. (1690). *Essay Concerning Human Understanding*. London: Basset.

Meillet, A. (1912). L'évolution des formes grammaticales. *Scientia*, 12:26, Milan. (Reprinted 1951, in *Linguistique historique et linguistique générale*, pp. 130–48. Paris: Klincksieck.)

Metcalf, G. J. (1974). The Indo-European hypothesis in the sixteenth and seventeenth centuries. In D. Hymes (ed.), *Studies in the History of Linguistics: Traditions and Paradigms* (pp. 233–57). Bloomington, IN: Indiana University Press.

Muller, J.-C. (1986 [1786]). Early stages of language comparison from Sassetti to Sir William Jones (1786). *Kratylos*, 31, 1–31.

Osthoff, H. and Brugmann, K. (1878). *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*. Leipzig: S. Hirzel.

Owens, J. (1988). *The Foundations of Grammar: An Introduction to Medieval Arabic Grammatical Theory*. Studies in the History of the Language Sciences, 45. Amsterdam: John Benjamins.

Paul, H. (1920 [1880]). *Prinzipien der Sprachgeschichte*, 5th edn. Tübingen: Max Niemeyer.

Pedersen, H. (1962 [1931]). *The Discovery of Language: Linguistic Science in the Nineteenth Century*. Bloomington, IN: Indiana University Press.

Percival, W. K. (1986). Renaissance linguistics: the old and the new. In T. Bynon and F. R. Palmer (eds), *Studies in the History of Western Linguistics, in Honour of R. H. Robins* (pp. 56–68). Cambridge: Cambridge University Press.

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: HarperCollins.

Rask, R. K. (1993 [1818]) *Investigations of the Origin of the Old Norse or Icelandic Language*, translated by Niels Ege. Copenhagen: The Linguistic Circle of Copenhagen. (Originally published as *Undersögelse om det gamle Nordiske eller Islandiske Sprogs Orindelse*. Copenhagen: Gyldendal.)

Reuchlin, J. (1506). *De rudimentis Hebraicis libri tres*. Pforzheim: Thomas Anselm. (Reprinted 1974, Hildesheim: Georg Olms.)

Robins, R. H. (1990). *A Short History of Linguistics*, 3rd edn. London: Longman.

Sajnovics, J. (1770). *Demonstratio idiorna Ungarorum et Lapponum idem esse*. Copenhagen: Typis Collegi Societatis Iesu. (2nd edn, Trnava (Tyrnau), Hungary, 1770; photolithic reproduction, Bloomington, IN: Indiana University Press; The Hague: Mouton, 1968.)

Sanctius (Brocensis), Franciscus [Francisco Sanchez (de las Brozas)]. (1587). *Minerva seu de causis linguae latinae*. Salamanca: Ioannes and Andreas Renaut fratres.

Sapir, E. (1921). *Language*. New York: Harcourt, Brace, and Co.

Saussure, F. de (1959 [1916]). *Course in General Linguistics*. New York: McGraw-Hill.

Schlegel, K. W. F. von. (1808). *Ueber die Sprache und Weisheit der Indier*. Heidelberg: Mohr und Zimmer.

Schleicher, A. (1861–2). *Compendium der vergleichenden Grammatik der indogerm schen Sprachen*, 2 vols. Weimar: Hermann Bhlau.

Schmidt, J. (1872). *Die Verwandtschaftsverhaltnisse der indogermanischen Sprachen*. Weiman.

Skinner, B. F. (1957). *Verbal Behavior*. New York: Appleton-Century-Crofts.

Staal, J. F. (1974). The origin and development of linguistics in India. In D. Hymes (ed.), *Studies in the History of Linguistics: Traditions and Paradigms* (pp. 63–74). Bloomington, IN: Indiana University Press.

Stankiewicz, E. (1972). *A Baudouin de Courtenay Anthology*. Bloomington, IN: Indiana University Press.

Téné, D. (1995). Hebrew linguistic tradition. In E. F. K. Koerner and R. E. Asher (eds), *Concise History of the Language Sciences: From the Sumerians to the Cognitivists* (pp. 21–8). Oxford: Pergamon.

Traugott, E. C. and Heine, B. (1991). Introduction. In E. C. Traugott and B. Heine (eds), *Approaches to Grammaticalization* (pp. 1–14). Typological Studies in Language, 19. Amsterdam: John Benjamins.

Verner, K. (1877). *Eine Ausnahme der ersten Lautverschiebung. Zeitschrift fur vergleichende Sprachforschung*, 23, 97–130. (Translated in W. Lehmann (ed.), *A Reader in 19th Century Historical*

Indo-European Linguistics. Bloomington, IN: Indiana University Press.)

Weinreich, U., Labov, W., and Herzog, M. I. (1968). Empirical foundations for a theory of language change. In W. Lehmann and Y. Malkiel (eds), *Directions for Historical Linguistics* (pp. 95–195). Austin, TX: University of Texas Press.

Winteler, J. (1876). *Die Kerenzer mundart des Kantons Glarus in ihrem Grundzugen dargestellt*. Leipzig and Heidelberg.

Wundt, W. M. (1900). *Volkerpsychologie II: Die Sprache*. Leipzig: Engelmann.

7 Generative Grammar

Rule Systems for Describing Sentence Structure

THOMAS WASOW

1 Introduction

1.1 "Grammar"

To most people, the word "grammar" suggests a normative enterprise, dictating what constitutes correct language use. For example, many educated English speakers would identify at least five supposed grammatical "errors" in the following sentence:

(1) Hopefully, we will be able to easily figure out who to talk to.

Yet native speakers of American English also certainly recognize that (1) would be an entirely acceptable and natural sounding sentence in ordinary discourse. Indeed, the supposedly "correct" alternative (2) would be an awkward and affected way of expressing the thought.

(2) I hope that we shall be able easily to figure out to whom to talk.

Modern grammarians have little use for this prescriptive conception of grammar. They are more interested in the knowledge of English that allows native speakers to judge (1) as fully acceptable and (2) as somewhat less natural. The prescriptions of traditional grammar are largely attempts to impose the speech patterns of one region, class, ethnicity, or generation on speakers belonging to other groups. They may shed light on some of the social functions of language, and hence may be of interest to sociolinguists, anthropologists, historians, and political scientists; but they tell us very little about the structure of language. While some linguists study prescriptive grammar, few, if any, engage in it.

Language is a natural phenomenon, constituting an essential component of every human society. Linguistics is concerned with studying languages and language in general, much as biology studies living species and life in general. From this scientific perspective, the norms of prescriptive grammar are to linguists as the American Kennel Club's breed standards are to biologists: arbitrary evaluative standards of no relevance to objective description.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

Linguists use the term "grammar," then, to refer to structural properties of language that have evolved naturally and that native speakers of the language have mastered without explicit instruction. These are largely properties of languages that are not even mentioned in traditional grammars, though some are addressed in foreign language instruction. They include facts about word order, for example, that *we*, *will*, and *be* in (1) must appear in that order, or else the sentence becomes unacceptable. They also include facts about the proper forms of words in particular contexts, for example, that replacing *figure* in (1) with *figured*, *figures*, or *figuring* makes the sentence unacceptable. Put in more technical jargon, "grammar" is taken by linguists to encompass syntax and morphosyntax. The term may also be construed more broadly to include principles relating linguistic forms to the meanings they express (semantics) and/or the sound patterns of languages (phonology).

1.2 "Generative"

The term "generative" is associated with the tradition of grammatical research initiated and inspired by the work of Noam Chomsky. This term is sometimes construed very narrowly to refer only to work directly derivative from Chomsky's. Here it will be used more broadly to refer to work generally within the Chomskyan tradition, irrespective of whether its formalism and terminology come directly from Chomsky.

Among Chomsky's most important insights is the observation (noted independently over a century earlier by the great German linguist Wilhelm von Humboldt) that there are infinitely many well-formed sentences in any natural language. This follows immediately from the fact that any limit one might impose on the length of sentences would be arbitrary: any supposedly longest English sentence S would be two words shorter than *I said S*, which is surely well-formed if S is. On the other hand, a grammar, conceived of as a description of a language, should be finite.

How can we give a finite description of something infinite? Inspired by earlier work in mathematical logic and the foundations of computer science, Chomsky answered this question by proposing that we think of grammars as devices that put pieces of sentences together according to precise rules, thereby "generating" well-formed sentences. If some of the grammar rules can apply to their own outputs (in technical jargon, if some rules are "recursive"), then it is possible for finite grammars to generate infinite languages.

To illustrate this, consider the following very simple (nonlinguistic) example. The ordinary Arabic numeral system used to represent positive integers has infinitely many well-formed expressions (one for each number) constructed out of ten symbols, namely, the digits *0* through *9*. We can write a simple grammar for the numerals denoting positive integers with the following rules:

- Each of the digits *1, 2, 3, 4, 5, 6, 7, 8,* and *9* is a numeral.
- If N is any numeral, then N*0* is a numeral.
- If N_1 and N_2 are arbitrary numerals, then N_1N_2 is a numeral.

One of many possible formalizations of this would be the following:

$N \rightarrow 1$	$N \rightarrow 5$	$N \rightarrow 9$
$N \rightarrow 2$	$N \rightarrow 6$	$N \rightarrow N0$
$N \rightarrow 3$	$N \rightarrow 7$	$N \rightarrow NN$
$N \rightarrow 4$	$N \rightarrow 8$	

Here N is the category of well-formed numerals, and the arrow can be interpreted to mean "may consist of." This little grammar generates the infinite "language" of numerals denoting positive integers, because it contains rules that are recursive (namely, the last two).

2 Tenets of Generative Grammar

Although the term "generative" originally characterized a conception of grammars as such recursive rule systems, the term is now used somewhat more generally. In particular, what distinguishes work in generative grammar is the goal of describing languages systematically, as opposed to the more anecdotal approach of traditional grammars. While it is impossible to give a precise definition of generative grammar, there are several tenets shared by the vast majority of generative grammarians. These are summarized in the following subsections.

2.1 Grammars should be descriptive, not prescriptive.

As discussed above, this proposition is generally accepted by modern linguists. Although it is not unique to generative grammarians, it is common to them.

2.2 Grammars should characterize competence, not performance.

Despite its anti-prescriptivism, generative grammar is not an attempt to describe all or only the actual utterances of native speakers. This is implicit in the claim that languages are infinite: it would have been safe to assume that no sentence over one million words long will ever be uttered. But this upper bound exists because of limits on human memory and patience, not because of any linguistically interesting facts. Moreover, because of speech errors of various kinds, people frequently produce utterances that are not well-formed sentences, even by the judgments of the speakers. To distinguish between the idealized infinite languages that generative grammarians seek to describe and the far messier output of actual speakers, Chomsky introduced the terminology "competence" vs. "performance."

One common property of generative grammar in all its varieties is the focus on characterizing linguistic competence. Many generative grammarians would also like to develop models of linguistic performance, but most believe that a competence theory will be a necessary component of such a model. Put slightly differently, it is widely accepted that explaining how a language is actually used will require understanding speakers' unconscious knowledge of that language.

2.3 Grammars should be fully explicit.

Traditional grammars presuppose some knowledge of the language under description and tend to focus on aspects of the language that are variable or have changed. Generative grammars are supposed to be precise rule systems that characterize the whole language, without relying on any prior knowledge of the language on the part of the reader. Many generative grammarians identify explicitness with formalization. Hence, the generative literature abounds with formalisms (though it is not always made clear how the formalisms are to be interpreted). Early work in generative grammar approached this goal of explicitness and formalization far more consistently than most recent work.

2.4 Linguistic analyses should be maximally general.

If two grammars cover the same range of data, but one requires two distinct rules where the second has only one, generative grammarians take this as evidence for the superiority of the second grammar.

A famous example of this mode of reasoning is due to Postal (1964). He noted that what are called "tag questions" in English require a kind of matching between the tag and the initial

portions of the main clause, as illustrated in (3). Following standard practice, asterisks are used to mark unacceptable strings.

(3) a. I have won, { haven't / *won't / *aren't } { I / *you / *we / *they } ?

b. You will win, { won't / *haven't / *aren't } { you / *we / *they } ?

Postal also observed that imperative sentences take only a restricted range of tags, though there is nothing overtly present in the initial portions of imperative sentences that the tags match.

(4) Close the door, { *haven't / won't / *aren't } { *I / you / *we / *they } ?

If we analyze imperative sentences as having an initial *you will* at some level of analysis, he reasoned, we could use a simple rule to generate tag questions on both declarative and imperative sentences. Such an analysis is said to "capture a generalization" – in this case, the generalization that tags on imperatives and declaratives are fundamentally alike. The desire to capture generalizations plays a very important role in the argumentation of generative grammar.

2.5 *The theory of grammar should make universal claims.*

To the extent possible, facts about individual languages should be derived from general principles that apply to all languages. Information stipulated in the grammars of particular languages should be kept to a minimum. This is motivated in part simply by standard scientific methodological considerations: more general hypotheses are both more parsimonious and more interesting than less general ones. But it is also motivated in part by psychological concerns – specifically, by Chomsky's "argument from the poverty of the stimulus," which will be discussed in the next subsection.

The focus on the development of a general theory of grammar – "universal grammar" (UG), as Chomsky dubbed it – is perhaps the most distinctive characteristic of the generative tradition. Although other linguistic traditions involve extensive crosslinguistic comparisons resulting in important hypotheses about universal properties of language (see Chapter 3 for a sample of such work), generative grammar approaches these issues in a distinctive way. Specifically, the universals of generative grammar tend to be formulated as rather abstract principles of grammatical organization that are not directly observable in the linguistic data. Rather, their discovery and testing typically involve a complex combination of empirical observations, methodological assumptions, and inferential processes. This is in sharp contrast with more observationally transparent universals like those of Greenberg (1963), and much subsequent work on language typology. Some examples of linguistic universals in the generative style will be provided in Section 4 below.

2.6 *Grammars should be psychologically relevant.*

Generative grammarians characteristically (but not universally – see, for example, Katz and Postal 1991) take their theories to be relevant to psychological questions. Chomsky has been particularly outspoken on this issue, asserting that "a particular generative grammar" is "a theory

concerned with the state of the mind/brain of the person who knows a particular language." (Chomsky 1986: 3).

More specifically, Chomsky has argued that a rich theory of universal grammar is necessary to account for the possibility of language acquisition. The most striking fact about human languages, he claims, is the gulf between knowledge and experience, observing that the following question, formulated by Bertrand Russell, is particularly applicable in the domain of language:

> How comes it that human beings, whose contacts with the world are brief and personal and limited, are nevertheless able to know as much as they do?

The fact that every normal human masters a language with little apparent effort or explicit instruction suggests that humans are genetically endowed with a "mental organ" specifically adapted to acquire languages of a particular kind. This is known as the "argument from the poverty of the stimulus."

While Chomsky has emphasized the issue of learnability, others have argued that work in generative grammar is relevant to psychology in other ways. For example, Bresnan (1978) argued that a generative grammar should be an integral component of a theory of language use – that is, of the mental processes involved in speaking and comprehension.

3 Common Formal Elements

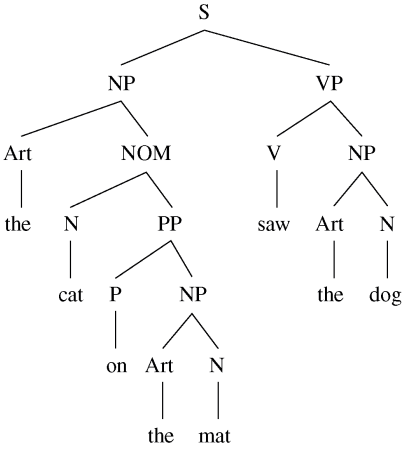
Since Chomsky's seminal work in the 1950s, many different theories of grammar have been articulated that fit the general characterization in the preceding sections. Almost all can be viewed as extensions of what is known as "context-free (phrase structure) grammar" (CFG).

3.1 Context-free grammar

CFG begins with the relatively uncontroversial assumption that words can be classified into categories, based on their morphological properties (that is, what changes in form they undergo through suffixation and the like), their distributional patterns (that is, what other words appear in their vicinity in sentences), and their meanings. The traditional categories of noun, verb, etc. (inherited from the grammatical studies of ancient Greece) are still quite generally employed, supplemented by a number of other categories, some of them idiosyncratic to particular theories.

A second generally accepted premise of CFG is that the words in sentences are grouped into phrases, which themselves are grouped together into larger phrases, and so on. It is common to represent the phrase structure of a sentence by means of a "tree diagram" like (5).

(5)



Phrases are identified by their distributional patterns and usually function as semantic units as well. Like words, phrases are generally classified into categories; the most widely used phrasal category labels – e.g., noun phrase (NP), verb phrase (VP), prepositional phrase (PP) – derive from the categories of words that appear in canonical instances of those phrases. These words are called the "lexical heads" (or sometimes just the "heads") of the phrases.

A CFG has two parts:

- A lexicon, consisting of a list of words, with their associated grammatical categories.
- A set of rules of the form A→Φ where A is a phrasal category, and Φ stands for any string of lexical and/or phrasal categories. The arrow is to be interpreted as meaning, roughly, "may consist of." These rules are called "phrase structure rules."

The left-hand side of each rule specifies a phrase type (including the sentence as a type of phrase), and the right-hand side gives a possible pattern for that type of phrase. Because phrasal categories can appear on the right-hand sides of rules, it is possible to have phrases embedded within other phrases. In fact, some types of phrases (such as NPs and PPs) can be embedded in other phrases of the same type, giving CFGs the recursive character needed to generate infinite languages.

A CFG normally has one or more phrasal categories that are designated as "initial symbols." These are the types of phrases that can stand alone as sentences in the language. Most simple CFGs have just one initial symbol, namely S. Any string of words that can be derived from one of the initial symbols by means of a sequence of applications of the rules of the grammar is generated by the grammar. The language a grammar generates is simply the collection of all of the sentences it generates.

In the 1950s and early 1960s, Chomsky, Postal, and others argued that simple CFGs lacked the descriptive power to account for all of the syntactic regularities of natural languages. Although some of those arguments have since been called into question, the conclusion remains generally accepted (see Savitch et al. 1987 for a collection of relevant articles).

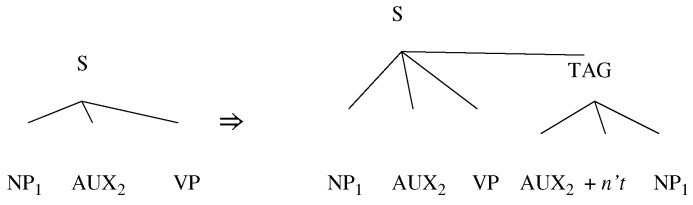
3.2 Transformational grammar

Chomsky's earliest work suggests that the shortcomings of CFG could be remedied by associating with each sentence of a natural language not just one tree but a sequence of trees. The initial tree in each sequence would be generated by a CFG (sometimes called the "base") and subsequent trees would be derived through a series of transformations – that is, rules that modified the trees in precisely specified ways.

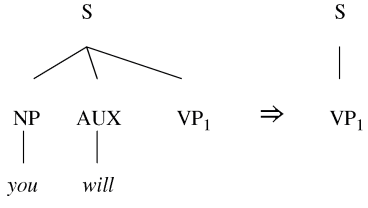
This can be illustrated with the phenomena of tag questions and imperatives described above (see Chapter 13 by Baker for further illustrations). Space limitations require substantial simplifications: only non-negative sentences with pronouns as subjects and auxiliary verbs will be considered here. A simple transformational grammar for these phenomena might include the base grammar in (6) and the transformations in (7) and (8). Parentheses are used to indicate that an element is optional – for example, the fourth rule in (6) says a VP may consist of a verb, with or without a following NP. In (7) and (8), '⇒' means "may be transformed into."

(6) A lexicon for English, plus:
S→NP AUX VP
NP→Pronoun
NP→(Art) N
VP→V (NP)

(7)



(8)



Example (7) takes as input trees for simple declarative sentences, and produces as outputs trees for the corresponding sentences with tags. It does this by copying the auxiliary verb, inserting a contracted negative, and copying the subject. Example (8) derives imperative sentences from declaratives starting with *you will* simply by deleting these two words (and the nodes right above them).

On this analysis, (4) is derived from the same base tree as *You will close the door*, by application of the two transformations, in the order given.

Early generative work was known as "transformational grammar," because the addition of transformations to CFG was seen as the crucial innovation. Throughout the history of generative grammar, transformational theories have had many advocates – always including Chomsky. Since the late 1970s, however, nontransformational alternatives have also been extensively developed.

3.3 Other enhancements to CFG

Several enhancements to simple CFG have been adopted in transformational and nontransformational generative theories alike. One of the earliest was the addition of a semantic component. It is evident that the acceptability of a sentence is influenced by what the intended meaning is, and it is often difficult to draw a sharp line between syntactic and semantic analyses. Consider, for example, the facts in (9).

(9) a. I excused myself.
b. *I excused me.
c. He excused himself. [*He* and *himself* must refer to the same person]
d. He excused him. [*He* and *him* must refer to different people].

The facts in (9a) and (9b) are manifestly about the distribution of the words *myself* and *me*. The contrast between (9c) and (9d) is evidently a semantic one. Yet there is clearly a single generalization covering both contrasts, namely, that in the configuration NP_1-V-NP_2, NP_2 can be a reflexive pronoun (that is, a form ending in *-self* or *-selves*) just in case it refers to the same individual as NP_1. This generalization will be developed in more detail below. For now, it can serve as an illustration of the role of semantics in grammar.

Another enhancement of CFG that has been generally adopted in generative grammar is the use of nonatomic category labels for words and phrases. For example, in the mini-grammar presented in (6), AUX and V are distinct categories, with no more in common than, say, N and V. But what this grammar calls AUX has traditionally been treated as a species of verb. This makes

sense when one considers the word *have*. In a sentence like *We have won*, (6) would treat it as an AUX (consider *We have won, haven't we?*); in a sentence like *We will have fun*, (6) must treat *have* as an instance of V (consider *We will have fun, won't we?*).

There are many more arguments for allowing words and phrases to be treated as the same in some respects but different in others. This is accomplished formally by replacing atomic category labels with more complex information structures. In particular, almost all varieties of generative grammar employ "feature structures" as category labels. Feature structures consist of pairings of features with values. A feature is simply a name for something used in classifying words or phrases; features are associated with multiple values, corresponding to properties of the words or phrases in question. For example, nouns can be subclassified into proper and common nouns, and into singular and plural nouns. Representing this with features would involve positing two features, say COMMON and NUMBER, each of which has two values (in English, at least). Then the two features could be used in representing the categories of some representative words as follows:

(10)

child	COMMON +	*London*	COMMON –	
	NUMBER sing		NUMBER sing	
children	COMMON +	*Alps*	COMMON –	
	NUMBER pl		NUMBER pl	

All of the feature structures in (10) might also have something like [POS noun] (where POS is for "part of speech").

Treating categories as bundles of features makes it possible to represent large numbers of grammatical categories quite compactly, since every different combination of features and values is a different category. This allows grammarians to make fine distinctions, while still permitting reference to large classes of expressions. Some form of decomposition of categories into features has consequently been adopted in almost every variety of generative grammar. So long as there are only a finite number of features, each of which has only a finite number of possible values, this decomposition does not fundamentally alter the descriptive power of CFG. It does, however, make it possible to capture generalizations across categories of words and phrases, as well as characterizing categories at more or less fine-grained levels.

Some theories have taken this process one step further, however, allowing the values of features to be feature structures themselves. This constitutes a more fundamental enhancement of CFGs, allowing a great deal of information to be encoded into the representations of grammatical categories. As will become evident below, this increased descriptive power makes possible interesting alternatives to certain widely accepted transformational analyses.

One of the advantages of decomposing categories into features is that it permits efficient reference to classes of categories. For example, one can refer to all singular nouns with the feature specification

POS	noun
NUMBER	sing

leaving other properties unspecified, including gender, case, and whether it is proper or common. This sort of "underspecification" is widely exploited in generative grammar.

One particularly influential case of underspecification is the suggestion by Chomsky (1970) that the phrase structure rules of languages could be reduced to a few very general schemas, with highly underspecified categories on both sides of the rules. This idea has been developed in many different ways, but has found its way into most theories of generative grammar. In its simplest version, it holds that all phrases should be viewed as projections of lexical heads and that phrases uniformly have three levels: the lexical head, an intermediate level, and the full

phrase. These are often designated as X, X′, and X″ (where X can stand for any combination of features). Then the principal phrase structure rules could be schematized as follows (where the superscripted asterisk is an abbreviation for zero or more occurrences of the immediately preceding symbol):

(11) $X'' \rightarrow (Y'')\ X'$ $\qquad\qquad$ $X' \rightarrow X\ Y''^{*}$

These rule schemas embody the claim that all phrases have the same basic structure, consisting of a lexical head, possibly followed by some other phrases (known as "complements"), and possibly preceded by a single phrase (known as the "specifier"). Variants of this idea go under the label "X-bar theory." Although there are many different versions of X-bar theory, schematizing the phrase structure rules through underspecification of the categories is common to many generative theories.

The rule schemas in (11), as stated, do not appear to be good candidates for universal grammar, because they stipulate particular orders of elements. But there are languages (such as Japanese) in which lexical heads consistently come at the ends of phrases, and others (such as Irish) in which lexical heads come at the beginnings of phrases. It has been proposed (e.g., by Gazdar and Pullum 1981) that the information about hierarchical structure and the information about left-to-right ordering of elements should be decoupled. That way, the schemas in (11) could be regarded as universal, up to the ordering of elements on the right-hand sides. This is another idea that has emerged in a number of different generative theories.

4 Some Phenomena Studied by Generative Grammarians

The literature of generative grammar is full of detailed examinations of myriad syntactic phenomena in a wide variety of languages. Most analyses depend on assumptions that are controversial. Nevertheless, the field has made numerous genuine discoveries. Although different schools of thought employ disparate formalisms and terminology, we know far more about the structure of language than we did in the 1950s, thanks to research in generative grammar. This section provides an overview of two areas in which generative grammarians have made clear progress.

4.1 Binding principles

The examples in (9) above illustrate that English has two different types of pronouns, namely reflexives (*-self/-selves* forms) and nonreflexives. While *myself* and *me* both refer to the speaker (as does *I*), the environments in which they can be used differ. In particular, consider the following contrasts:

(12) a. *I support me.
b. I support myself.
c. They support me.
d. *They support myself.

(13) a. I don't expect them to support me.
b. *I don't expect them to support myself.
c. *They don't expect me to support me.
d. They don't expect me to support myself.

The following two generalizations (known as "binding principles") roughly summarize the distributional difference between the two types of pronouns:

- A. A reflexive pronoun must have a local antecedent.
- B. A nonreflexive pronoun may not have a local antecedent.

For present purposes, "antecedent" can be taken to mean a preceding NP with the same reference. The term "local" is meant to convey the observation that the antecedent of a reflexive should not be too far away. However, giving a precise definition of "local" for these principles is not a trivial problem, as evidenced by examples like the following:

(14) a. The house has a fence around (it / *itself)

b. We wound the rope around (it / itself) [it ≠ the rope]

c. I wrapped the blanket around (me / myself)

These examples show that locality cannot be measured simply in terms of number of words or phrases intervening between the pronoun and its antecedent, for the three examples all have the same number of words and phrases.

There is a rich literature dedicated to working out the details of the basic ideas in principles A and B above. These details need not concern us here. What is of interest is that English is by no means unique in having these two different kinds of pronouns. Indeed, a great many languages have parallel sets of pronouns that differ in just this way: one kind requires local antecedents and the other prohibits local antecedents. Just what counts as "local" (an issue we will return to) exhibits some crosslinguistic variation, but the similarity is more striking than the difference. There is no a priori reason to expect languages to have more than one kind of pronoun, yet something like the principles above hold in language after language.

Notice, incidentally, that the binding principles interact in an interesting way with the analysis of imperatives suggested in Section 3.2. Assuming that the principles are applied prior to the deletion of *you*, the principles correctly predict the following:

(15) a. Protect { *myself / yourself / *himself } !

b. Protect { me / *you / him } !

This provides further evidence that imperatives should be treated as having second-person subjects at some level of analysis.

4.2 Filler-gap dependencies

Context-free grammars provide a formal mechanism for expressing relationships between elements (words or phrases) that are close to one another in a sentence. But many languages have constructions involving dependencies between elements that may be far apart. An example of this in English is what are known as "*wh*-questions" – that is, questions requiring more than a yes-or-no answer, and hence containing one of the "*wh*-words" (*who, what, where*, etc.).

To illustrate this, consider the examples in (16).

(16) a. Pat relies { on / upon / *of / *to } a student.

b. { On / Upon / *Of / *To } which student does Pat rely?

c. { On / Upon / *Of / *To } which student does Chris say we think Pat relies?

(16a) shows that the verb *rely* requires a prepositional phrase complement beginning with *on* or *upon*; (16b) shows that, in a *wh*-question, although this phrase comes at the beginning of the sentence, rather than after the verb, the same restriction on the choice of prepositions is maintained; (16c) illustrates that this dependency between the verb and preposition holds even when lots of other material is inserted between them. In fact, there is no limit to the amount of additional text that can intervene.

Similarly, the dependency between verb form and the number (singular or plural) of its subject is preserved, even when the subject is a *wh*-phrase that is far away in the string of words.

(17) a. The (teacher / *teachers) dislikes one student.

b. The (*teacher / teachers) dislike one student.

c. Which (*teacher / teachers) would the parents all claim dislike one student?

d. Which (teacher / *teachers) would the parents all claim dislikes one student?

More generally, *wh*-phrases in such questions behave in some ways as though they were in a different position from where they actually occur. Dependencies like preposition selection or verb agreement, which are normally local, can hold between *wh*-phrases and elements far away in the sentence. This can be further demonstrated with the binding principles:

(18) Which dog do you think we saw scratch { *yourself / you / itself / it [it ≠ which dog] }

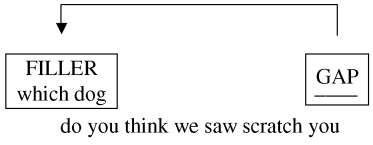
On the surface, *which dog* does not look like the required local antecedent for *itself*, because of the intervening material *do you think we saw*. Moreover, *you* cannot serve as the antecedent for a reflexive object of *scratch*, even though it is closer to the object position. The binding pattern here is just what principles A and B would predict if *which dog* were in the subject position of *scratch*.

A very natural way to account for such relationships in a transformational grammar is to posit a rule that moves *wh*-phrases to the front of the sentence. Then the *wh*-phrases in (16), (17), and (18) can be generated initially in a position close to the relevant verb or reflexive, and the dependencies can be licensed locally, prior to movement.

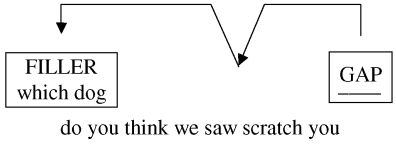
With such a treatment of *wh*-questions and similar constructions, a question that naturally arises is whether the displaced elements (often referred to as "fillers") move from their initial positions (known as "gaps") to their final positions in one fell swoop or by means of a sequence of smaller movements. That is, in an example like (18), does the filler *which dog*, move from the gap position adjacent to *scratch* in one long movement, as in (19a), or in several smaller movements, as in (19b).

(19)

(a)



(b)



The issue can be formulated in a more theory-neutral way by asking whether the relationship between a gap and its filler is a direct one, or is instead mediated by intervening material. This was a hotly debated topic within generative grammar in the 1970s (sometimes labeled the "swooping vs. looping" controversy). A real measure of progress in the field is that this debate has been definitively settled in favor of "looping." All generative grammarians now recognize that long-distance filler-gap dependencies are mediated by the intervening material.

The key evidence for this comes from languages that require some sort of marking of clauses that intervene between fillers and gaps. Quite a number of such cases have been discovered, from a wide range of language families (see Zaenen 1983 for presentation of a few). Exactly where in the intervening clauses the marking occurs, and which form it takes, varies from language to language (though there seem to be some regularities).

A clear and relatively simple example is the relative clause construction[1] in Irish. Irish relative clauses, like those in English, immediately follow the noun they modify, and must contain a gap. The filler for the gap is the noun the clause modifies. Now consider the following examples (adapted from McCloskey 1979):

(20) a. Mheas mé gur thuig mé an t-úrscéal
thought I that understood I the novel
'I thought that I understood the novel.'

b. an t-úrscéal a mheas mé a thuig mé
the novel that thought I that understood I
'the novel that I thought I understood'

c. Shíl mé go mbeadh sé ann
thought I that would-be he there
'I thought that he would be there.'

d. an fear a shíl mé a bheadh ann
the man that thought I that would-be there
'the man that I thought would be there'

e. Dúirt mé gur shíl mé go mbeadh sé ann
said I that thought I that would-be he there
'I said that I thought that he would be there.'

f. an fear a dúirt mé a shíl mé a bheadh ann
the man that said I that thought I that would-be there
'the man that I said that I thought would be there'

g. an fear a shíl go mbeadh sé ann
the man that thought that would-be he there
'the man that thought he would be there'

h. an fear a dúirt sé a shíl go mbeadh sé ann
the man that said he that thought that would-be he there
'the man that said he thought he would be there'

i. an fear a dúirt gur shíl sé go mbeadh sé ann
the man that said that thought he that would-be he there
'the man that said he thought he would be there'

Underlining indicates the regions of these sentences that are between gaps and their fillers. That is, the word immediately preceding each underlined piece is a filler for a gap located immediately after the underlining. Now look at the words that have been translated as *that*. Where there is no underlining, the Irish equivalent of *that* is either *go* or *gur* (the difference between them is not relevant to the present discussion). But wherever *that* translates an underlined word, the word it translates is *a*. These words are known as "complementizers" (see Chapter 13 by Baker for more discussion of complementizers), because they introduce clausal complements to verbs like *mheas* ('thought'), *shíl* (also translated as 'thought'), and *dúirt* ('said'). Examples like those in (20) indicate that Irish employs different complementizers in the region between a filler and a gap than elsewhere.

Modern transformational analyses of filler-gap relationships posit movement through a series of intermediate positions. This fits well with the Irish data, if the complementizer *a* serves as a special gateway through which long-distance movements must pass.

4.3 Island constraints

The notion of gateways for filler-gap dependencies has also been useful in discussions of another much-studied set of phenomena. Although there is no bound on the distance between fillers and gaps, there are a number of constraints on the relative positions in which fillers and their corresponding gaps may appear. These are known as "island constraints," following Ross (1967).

One such restriction on filler-gap dependencies is that the gap may not be in a relative clause if the filler is outside of it. Thus, for example, *wh*-phrases in English questions cannot fill gaps inside of relative clauses, as illustrated in (21). The relative clauses are enclosed in square brackets, and the gap positions are marked "___".

(21) a. *Which dog did you criticize the person [who kicked ___]?
b. *How many sources does the prosecutor have evidence [which was confirmed by ___]?
c. *Who did everyone envy the writer [whose book claimed ___ was the real Deep Throat]?

If a *wh*-phrase has to pass through intervening complementizer positions on its way from its initial positions (the gap) to its surface (filler) position, then it seems natural to block examples like (21) on the grounds that the relative clauses already have *wh*-phrases (*who*, *which*, and *whose book*) in their complementizer positions. Such an analysis would also rule out gaps internal to embedded questions, as in (22).

(22) a. *Which dog did they ask [who you said kicked ___]?
b. *How many sources does the defense wonder [why the prosecutor asked for ___]?
c. *Who did everyone inquire [whose book claimed ___ really wrote Shakespeare's plays]?

Not all island constraints are covered by this. For example, a gap cannot be in a coordinate conjoined structure not containing its filler, unless all conjuncts have gaps filled by the same filler:

(23) a. *What did they [buy ___ and forget their credit card at the store]?
b. What did they [buy ___ and forget ___ at the store]?

A great deal of research has gone into island constraints: classifying them, checking their cross-linguistic variation, and, most extensively, seeking explanations for them. The question of explaining island constraints will be addressed again below.

5 Varieties of Generative Grammar

As noted earlier, generative grammar is not so much a theory as a family of theories, or a school of thought. The preceding sections have focused on common elements: shared assumptions and goals, widely used formal devices, and generally accepted empirical results. (For convenience, the idiom of transformational grammar has been employed in the descriptions of tag questions, imperatives, and filler-gap dependencies, but the discussion in Section 5.2 below shows that this was not essential.) This section explores some of the ways in which generative theories differ from one another. There are too many such theories to provide a comprehensive survey (see Sag, Wasow, and Bender 2003, Appendix B for brief overviews of various theories of grammar), but the following sections characterize some of the major divisions, beginning with a brief description of the historical development of transformational grammar.

5.1 Transformational theories

Transformational grammar has evolved considerably over the decades (see Newmeyer 1986). The earliest work (Chomsky 1957) was concerned largely with showing the inadequacy of context-free grammar for the analysis of natural languages, and with providing precise, explicit transformational descriptions of particular phenomena (largely from English). In the 1960s, transformational grammarians began paying more attention to the relationship between syntax and semantics, leading to heated debates over the best way to incorporate a semantic component into transformational theory. At the same time, the emphasis turned away from providing rule systems in careful detail to exploring the wider implications of transformational analyses. This was when questions about Universal Grammar and the relevance of linguistic theory to psychology came to the fore (Chomsky 1965). Since the early 1970s, the primary focus of

transformationalists has been on developing a highly restrictive theory of grammar – that is, one that narrowly constrains what kinds of descriptions are possible (Chomsky 1981). The goal of this enterprise, as articulated by Chomsky, is to account for language learnability by making the theory so restrictive that a descriptively accurate grammar of any language can be inferred on the basis of the kind of data available to a young child.

As the goals and style of transformational grammar have evolved over the years, the technical details have changed as well – many almost beyond recognition. Through all these changes, however, this line of research has maintained the idea that sentences are derived by means of a sequence of operations that deform tree structures in prescribed ways. Inherent in this conception is a directionality: derivations proceed from underlying structures to surface forms. This directionality found its way into analyses sketched in this chapter wherever one rule or principle was said to operate "before" another. Examples are the treatment of imperative tags, in which the tag formation transformation had to operate before the imperative rule, and in the account of island constraints in terms of one *wh*-phrase getting into the complementizer position before another one needed to move there.

Many linguists find this sort of talk troublesome. Grammars are supposed to be characterizations of linguistic competence – that is, the knowledge of language that underlies both speaking and understanding. Speaking involves articulating thoughts – going from meanings to sounds; understanding involves extracting meanings from sounds. So, in an intuitive sense, these processes operate in opposite directions. The knowledge of language that is common to both should be process-neutral and hence nondirectional. It is possible to regard the talk of operations and directions as strictly metaphorical, a move that has sometimes been advocated. But translating from procedural formulations into more static ones is not always straightforward.

The problem is not just that readers tend improperly to read some psychological significance into the directionality inherent in transformational derivations (though this tendency certainly exists) but that psycholinguists and computational linguists who have tried to use transformational grammars as components in models of language use have found that transformational derivations are typically not easily reversible. Precisely worked-out systems to parse sentences – whether they are intended as models of human performance or as parts of computer systems for understanding languages – have almost never incorporated the transformational analyses proposed by theoretical linguists. These analyses do not lend themselves to being used in going from the surface form of a sentence to its meaning. Moreover, as noted by Fodor, Bever, and Garrett (1974), psycholinguists have been largely unable to find behavioral evidence for the psychological reality of the intermediate stages of transformational derivations. While the nature of the intermediate stages posited by transformational grammarians has changed radically since Fodor et al. made that observation, the observation itself remains accurate.

5.2 *Nontransformational analyses*

A variety of alternatives to transformational grammar have been developed (see, e.g., Gazdar et al. 1985; Bresnan 2001; Steedman 1996; Pollard and Sag 1994). Some grammatical theories have questioned the basic conception of phrase structure embodied in tree diagrams (e.g., Hudson 1984), but most are less radical departures. Instead, they build on context-free grammar, providing enhancements designed for the description of natural languages. This section offers a sample of what such descriptions are like by revisiting some of the phenomena discussed in earlier sections.

Consider first the imperative construction. Imperatives behave as though they had a second-person subject (i.e., *you*), based on evidence from tags and reflexives; but no subject appears in imperative sentences. The transformational analysis offered above posits two distinct trees for imperative sentences, one with a subject and one without. An alternative approach is to posit a single tree without an overt subject phrase, but with the information necessary to get the facts about tags and reflexives right.

Suppose that the category of a word is a complex feature structure (see Section 3.3) that contains within it at least the following: (i) information about what other kinds of elements it can appear with; (ii) information about its semantics; and (iii) information about how the syntactic information in (i) is linked to the semantic information in (ii). For example, the lexical entry for the verb *protects* should indicate (i) that it requires a third-person singular NP subject and an NP object; (ii) that it denotes the protection relation; and (iii) the roles played in that relation by the NPs' referents, namely, that the referent of the subject protects the referent of the object. One possible formalization of this information is the following:[2]

(24)

protects	
POS	verb
FORM	present-tense
SUBJ	NP_1 PER 3rd NUM sing
OBJ	NP_2
SEM	protect < NP_1 , NP_2 >

In most cases, the arguments of the semantic relation (that is, the elements between the angle brackets) are linked one-to-one to the syntactic arguments, such as the subject and object. That is the case in (24). In imperatives and some other constructions, however, there may be a mismatch. So, for example, the lexical entry for the imperative use of the verb *protect* might be something like (25).

(25)

protect	
POS	verb
FORM	imperative
SUBJ	none
OBJ	NP_2
SEM	protect < [NP_1 PER 2nd] , NP_2 >

This representation incorporates both the information that imperative *protect* has a second-person argument and that it has no subject. Further, the second-person argument is the one that plays the protector role in the semantics.

Now, in order to get facts like (15) right, it is necessary to interpret the binding principles as making reference to semantic argument structures. That is, the term "local" in the binding principles, which was left undefined in the earlier discussion, can now be taken to mean "in the argument structure of the same predicate." Thus, Principle A now says that a reflexive pronoun must have an antecedent that is an argument of the same predicate.

This characterization of locality makes an interesting new prediction: a reflexive pronoun in object position may not have an antecedent that is only part of the subject. That is, examples like (26) are correctly ruled out.

(26) a. *Your mother protects yourself.
b. *A picture of them upset themselves.

A definition of "local" in terms of simple proximity (based on either word strings or trees) would very likely not cover (26).

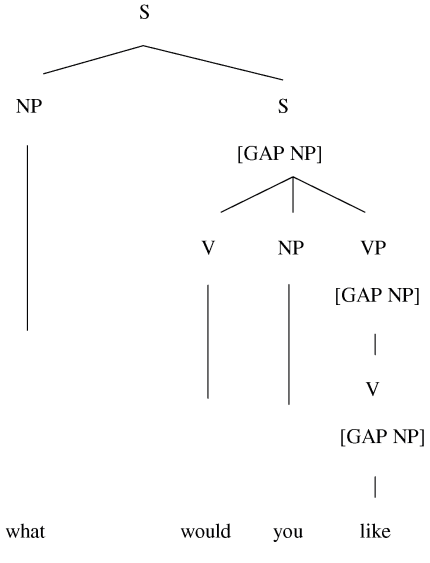
Filler-gap dependencies can be handled in a way that is at least partially analogous. A feature – call it GAP – taking another feature structure as its value can encode what is displaced. This information is represented on every node in the tree between the position of the gap and that of the filler. For example, in a sentence like *What would you like?* the category of *like* would include the information that it has no object, but that it has a GAP value that is linked to the second semantic argument of *like*, as in (27a). This GAP information would be shared by the VP and S nodes above *like* in the tree, as in (27b).

(27)

(a)

like	
POS	verb
FORM	infinitive
SUBJ	NP_1
OBJ	none
GAP	NP_2
SEM	like < NP_1 , NP_2 >

(b)



The phrase structure rule licensing the top part of this tree (where the gap is filled) must specify that the features on the NP must match those in the value of the GAP feature. The rule, then, is something like (28), where the identity of the subscripts is intended to indicate identity of all features.

(28) $S \rightarrow X_1''$ S[GAP X_1'']

Informally, what (28) says is that a clause containing a gap may be combined with a phrase of the appropriate type on its left to form a complete sentence (where the appropriate type of phrase is one that has the properties of the missing element that are encoded in the GAP value).

In addition, a principle is required that will guarantee that GAP values are shared between a node and the one immediately above it in a tree, except where rule (28) fills the gap. A GAP value on a node says that there is a gap somewhere within that phrase, and the filler for that gap is outside the phrase; the value of the GAP feature gives the syntactic details of the displaced element.

Many details have been left out of this account, but this minimal sketch is enough to address some of the phenomena discussed in Sections 4.2 and 4.3. First of all, the fact that local requirements can be satisfied by a distant filler follows from the fact that the filler must share all its features with the GAP value. Any local requirements on the GAP value must therefore be met by the filler. Recall, for example, (18). If a reflexive pronoun appears as the object of *scratch*, then Principle A requires an antecedent that is also an argument of the *scratch* relation. The feature structure for *scratch* in this sentence identifies the GAP value with the first argument of the *scratch* relation, and this GAP value must match the features of the filler, *which dog*. Hence, *which dog*, but not *you*, can be the antecedent of a reflexive in (18), despite their positions in the sentence.

Turning now to the swooping vs. looping controversy, it is evident that no such issue arises in this nontransformational analysis. The information about gaps that must be available at the position of the filler is transmitted through the intervening structure. Hence, the presence of a gap in a phrase is necessarily encoded in the category of the phrase. Phenomena like the Irish data in (20) are easy to account for: the choice of complementizer differs depending on whether the clause introduced has a GAP value.

Similarly, island constraints can be straightforwardly formulated in terms of the GAP feature. In fact, if GAP is formulated as suggested above, the island constraints discussed here are almost automatic consequences. Relative clauses and embedded questions are constructions that involve filler-gap dependencies. As long as GAP can have only one value, this makes it impossible to introduce a second gap inside one of these constructions. For example, in an embedded question like *who you said* ___ *kicked that dog* in (29), rule (28) licenses the combination of the filler *who* with the S[GAP NP] *you said* ___ *kicked that dog*.

(29) They asked who you said kicked that dog.

If one tried to question the object of *kicked*, yielding (22a), the phrase *you said kicked* would need to have two different GAP values, one corresponding to the subject of *kicked* and the other corresponding to its object.[3]

The facts illustrated in (23) – that filler-gap dependencies in coordinate constructions (i.e., phrases conjoined by *and* or *or*) are impossible unless they involve all conjuncts – are natural consequences of the analysis in terms of GAP. Coordinate conjuncts must share most syntactic features. For example, words or phrases with different grammatical categories cannot usually be conjoined, as in (30).

(30) a. *Everyone wishes for comfort and happy. [N *and* Adj]
b. *Pat became famous and in great demand. [Adj *and* PP]

Likewise, a VP whose FORM value is present-tense cannot be conjoined with one whose FORM is infinitive, as in (31).

(31) *Chris eats snails and drink wine.

If GAP is one of those features that must be identical across conjuncts, then facts like (23) are an immediate consequence. In a coordinate structure, either all conjuncts have an empty GAP value, or they all have the same nonempty GAP value. That is, they either all are gap free, or they all have a gap with the same filler.

6 The Future of Generative Grammar

Despite the variety of generative theories of grammar that have been put forward, the field has been dominated throughout its history by the work of one individual, Noam Chomsky. He was its founder; he has been its most prolific innovator; and the mainstream of generative research has always followed his lead. Even the proponents of alternative theories (such as the nontransformational approach sketched in the subsection 5.2) have usually taken work of Chomsky's as the point of departure for their proposals.

In the early years of generative grammar, the field was constituted largely by Chomsky and his students and collaborators. Over the decades, however, the number of generative grammarians grew exponentially. Under these circumstances, it is remarkable that Chomsky retained his dominant position for so long. But there are indications that this may be changing.

Given a saturated academic job market, increasing numbers of linguists are seeking employment in industry. This puts pressure on the field to give more attention to potential applications of its theories. The most obvious type of application for work in generative grammar would be in the development of natural language technologies – that is, computer programs that deal with human languages, e.g., doing machine translation, information retrieval from text files, summarization of texts, and the like. To the extent that such applications have motivated theoretical work, considerations of computational tractability have played an important role in theory construction. Likewise, such applications call for looking at how people actually use language, rather than focusing exclusively on what is grammatically possible. The investigation of actual usage data is greatly facilitated by the availability of large online text files, which can be sampled and analyzed with computational tools that did not exist until quite recently. This is already having a noticeable effect on the sorts of data used by generative grammarians in their theoretical arguments.

Along with the use of naturally occurring data, generative grammarians are paying increasing attention to their methods of eliciting judgments. Rather than relying on the intuitions of the investigator, more and more theoretical work employs laboratory methods to elicit judgments from substantial numbers of speakers, none of whom know what hypothesis is being tested. This has led to a blurring of the boundary between generative research and psycholinguistics, and more sophisticated experimental methods, such as reaction time measurement, eye tracking, and brain imaging, have been brought to bear on questions of syntactic structure.

In order to analyze these new sources of data, generative grammarians have needed to start using statistical tools in their work. This, in turn, has led to increased exploration of probabilistic models of language structure, acquisition, and use. Whether such work still falls under the umbrella of generative grammar is debatable, but it is clear that many of the practitioners of this new empiricist direction in linguistics began their careers as mainstream generativists. Hence, they are bringing new methods and new kinds of data to bear on some of the kinds of questions regarding linguistic structure that have been the focus of generative grammar.

These potential changes should not be worrisome. The history of generative grammar is one of numerous upheavals, as Chomsky has repeatedly modified the foundations of the theory. These upheavals have been accompanied by vigorous debates and lively competition from alternative frameworks. The result has been – and promises to continue to be – a robust line of research that has greatly enriched our understanding of human linguistic abilities.

SUGGESTIONS FOR FURTHER READING

Borsley, R. (1999). *Syntactic Theory: A Unified Approach*, 2nd edn. London: Hodder Education Group.
Carnie, A. (2006). *Syntax: A Generative Introduction*, 2nd edn. Oxford: Wiley-Blackwell.
Chomsky, N. (2006). *Language and Mind*, 3rd edn. Cambridge: Cambridge University Press.
Culicover, P. and Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.
Huddleston, R. and Pullum, G. K. (2005). *A Student's Introduction to English Grammar*. Cambridge: Cambridge University Press.
Lasnik, H. and Uriagereka J., with C. Boeckx (2005). *A Course in Minimalist Syntax*. Malden, MA: Blackwell.

RELEVANT JOURNALS AND SOCIETIES

Journals
Language
Linguistic Inquiry
Natural Language and Linguistic Theory
Syntax
Journal of Linguistics
Lingua

Societies
Linguistic Society of America
Linguistics Association of Great Britain
Deutsche Gesellschaft für Sprachwissenschaft

EMERGING TRENDS AND RESEARCH QUESTIONS

1. How does grammar interact with other cognitive faculties to produce observable linguistic behavior?
2. What is the biological basis for grammatical competence, both at the neurophysiological level and at the genetic level?
3. What role does grammar have to play in the development of robust language technologies?
4. How reliable are past results based on introspection? Do they need to be checked against usage data or controlled psycholinguistic experiments?
5. To what extent can the formal properties of language be explained in terms of the communicative functions of language?
6. What, if any, grammatical properties are universal to all human languages?

NOTES

1 Relative clauses are noun (or noun phrase) modifiers, such as the bracketed portion of the following:
(i) The student [that you rely on] isn't here yet.

2 This representation glosses over a great deal, including how the formalism is to be interpreted. Italics have been used in place of what should probably be a phonological representation, and underlining is used to designate a semantic relation, with the arguments in the relation listed immediately following, enclosed in angle brackets. The information in (24) also needs to be augmented by characterizations of subject and object in terms of tree configurations, but this is fairly straightforward, at least for English.

3 Both this explanation for these island constraints and the transformational one based on the idea of a blocked gateway rely on the presence of a filler-gap dependency in the embedded structure to establish its status as an island. This seems plausible for English, since overlapping filler-gap dependencies are not in general possible. Hence, in questions with multiple *wh*-words, only one can be a filler:
(i) What did Pat give to whom?
(ii) To whom did Pat give what?
(iii) *What to whom did Pat give?
(iv) *To whom what did Pat give?
But there are other languages (e.g., Polish and Greek) that permit analogues to (iii) and (iv). In those languages, GAP would need to be allowed to take multiple values (or a value that is a list of feature structures). Unfortunately, the correlation between restrictions on overlapping filler-gap dependencies and island constraints is not perfect, so these matters remain issues of ongoing research.

REFERENCES

Bresnan, J. (1978). A realistic transformational grammar. In M. Halle, J. Bresnan, and G. A. Miller (eds), *Linguistic Theory and Psychological Reality*. Cambridge, MA: MIT Press.

Bresnan, J. (2001). *Lexical Functional Syntax*. Oxford: Basil Blackwell.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1970). Remarks on nominalization. In R. A. Jacobs and P. S. Rosenbaum, *Readings in English Transformational Grammar*. Waltham, MA: Ginn-Blaisdell.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Fodor, J. A., Bever, T., and Garrett, M. (1974). *The Psychology of Language*. New York: McGraw-Hill.

Gazdar, G., Klein, E., Pullum, G. K., and Sag, I. A. (1985). *Generalized Phrase Structure Grammar*. Cambridge, MA: Harvard University Press and Oxford: Basil Blackwell.

Gazdar, G. and Pullum, G. K. (1981). Subcategorization, constituent order, and the notion 'head.' In M. Moortgat, H. van der Hulst, and T. Hoekstra (eds), *The Scope of Lexical Rules*. Dordrecht: Foris.

Greenberg, J. (1963). *Universals of Language* . Cambridge, MA: MIT Press.

Hudson, R. A. (1984). *Word Grammar*. Oxford: Basil Blackwell.

Katz, J. J. and Postal, P. M. (1991). Realism versus conceptualism in linguistics, *Linguistics and Philosophy*, 14, 515–54.

McCloskey, J. (1979). *Transformational Syntax and Model Theoretic Semantics*. Dordrecht: Reidel.

Newmeyer, F. J. (1986). *Linguistic theory in America, Second Edition*. Orlando, FL: Academic Press.

Pollard, C., and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press.

Postal, P. (1964). Underlying and superficial linguistic structure. *Harvard Educational Review*, 34, 246–66.

Ross, J. R. (1967). *Constraints on Variables in Syntax*. PhD dissertation, MIT. (Published as *Infinite Syntax!* (1986), Norwood, NJ: Ablex).

Sag, I. A., Wasow, T., and Bender, E. (2003). *Syntactic Theory: A Formal Introduction, 2nd edn*. Stanford, CA: CSLI Publications.

Savitch, W. J., Bach, E., Marsh, W., and Safran-Naveh, G. (1987). *The Formal Complexity of Natural Language*. Dordrecht: D. Reidel.

Steedman, M. (1996). *Surface Structure and Interpretation*. Cambridge, MA: MIT Press.

Zaenen, A. (1983). On Syntactic Binding. *Linguistic Inquiry*, 14, 469–504.

8 Functional Linguistics

Communicative Functions and Language Structure

ROBERT D. VAN VALIN, JR.

1 Introduction

If one were to take an informal survey among nonlinguists regarding the primary function of human language, the overwhelmingly most common answer would be, "language is used for communication." This is the commonsense view of what language is for. It might, therefore, come as surprise to many people that some of the most prominent linguists in the field reject this view and that many others hold that the fact that language may be used for communication is largely, if not completely, *ir*relevant to its study and analysis. Chomsky, for example, maintains that "human language is a system for free expression of thought, essentially independent of stimulus control, need-satisfaction or instrumental purpose" (1980: 239), and rejects the idea that communication is a necessary or even significant function of language as a "vulgar, instrumental" view of language (1975: 56–7; 1980: 229–30). Not all linguists share Chomsky's view, however, and many are strongly committed to a view of language which takes its role in communication as central to its study and analysis; they are not the majority in the field at present. Such linguists are referred to as *functionalists*, and the general term applied to this approach is *functional linguistics*. Within contemporary linguistics there is an opposition between functionalists, on the one hand, and *formalists*, on the other, formalists being those linguists who are in substantial agreement with Chomsky's position. As we will see later, this distinction has evolved into a more subtle and complex opposition than it might seem at first glance.

The goal of this chapter is to introduce the reader to the basic ideas of functional linguistics and to give an account of how the ideas that today constitute functional linguistics arose. It will also be explained how many professional linguists came to adopt a view of language which is so strikingly at odds with the view held by nonlinguists.[1] In Section 2, a number of terms and distinctions will be introduced that are relevant to elucidating functionalist and formalist approaches to the study of language. In Section 3, a brief history of twentieth-century linguistics will be given and the development of the relevant ideas about language structure and function will be sketched. In the final sections, contemporary functional linguistics will be characterized and contrasted with formal linguistics, to see how genuine the opposition really is.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

2 Communicative Functions of Language

What does the proposition "the primary function of language is communication" actually mean? What are the communicative functions of language? Many traditional accounts portray communication as being the conveying of propositions from the mind of one interlocutor to the mind of one or more other interlocutors, and the propositions are about some state of affairs, real or imagined. In the linguistic depiction of states of affairs, reference is made to entities in the states of affairs, and predications are made about actions involving the entities or relations among the entities in them. In this way speakers construct linguistic representations of situations, as in (1).

(1) The boy ate the bread in the kitchen.

There are three referring elements and one predicating element in (1): *the boy* [referring to one participant in the event] *ate* [predicating an action of the boy] *the bread* [referring to the second participant] *in the kitchen* [referring to the location where the event took place]. Hence reference and predication are often taken to be the fundamental communicative functions of language.

But language is used for much more than representing states of affairs. It is used in all kinds of verbal social interactions: asking questions, giving commands, making promises, expressing wishes, etc. These different uses are known as speech acts (Searle 1969). Foley and Van Valin (1984) emphasize the social nature of language use and stress that speaking is a kind of social activity:

> Communication is often construed in a narrow sense to mean 'conveying propositional information from one person to another', and within such a view linguistic behavior consists primarily of referring and predicating about situations in the world, all other types of verbal behavior, e.g., asking questions or giving commands, being derivative of it. Silverstein (1976, 1977, [1987]) has cogently argued that such a view is fundamentally mistaken and that referring-and-predicating is only one of the many socially constituted functions of language and not a privileged one at that… Thus the assumption that language is a system of communication treats language as a crucial component of human social interaction and takes linguistic behavior, e.g., asserting, asking, promising, commanding, wishing and requesting, and the larger-scale speech activities which they constitute, to be social behavior. (Foley and Van Valin 1984: 8)

It should be noted that the claim that the primary function of language is communication does not entail the view that all uses of language are necessarily communicative. Foley and Van Valin continue:

> There may well be instances of verbal behavior which are non-communicative, but this in no way undermines the fundamental functionalist tenet that an understanding of language structure requires an understanding of the functions language can serve… This position is analogous to claiming that in order to understand the structure of hammers it is necessary to know that they are used primarily for driving nails, even though they may also be employed as doorstops or paperweights or for tapping the ashes out of a pipe. Indeed, it would be difficult to account for the fact that the head of a hammer is always heavy metal and the handle wood or plastic and never vice versa, if one ignores its primary function, since a hammer could easily be a doorstop, paperweight or pipetapper with a plastic head and metal handle. Languages are much more complex than hammers, both structurally and functionally, but in both cases one cannot understand form independent of function. (Foley and Van Valin 1984: 8–9).

Thus, the function of conveying propositional information, i.e., linguistic depictions of states of affairs, is but one of many communicative functions that language has.

All of these different functions may have structural ramifications in languages. As a simple example, let's look at how two different languages express assertions (statements), interrogatives

(questions), and imperatives (commands). In English, each of these requires a different syntactic structure. This is illustrated in (2).

(2) a. The boy is eating the bread. Statement
b. Is the boy eating the bread? Question
c. Eat the bread! Command

In a statement, the subject precedes the tensed verb, be it an auxiliary verb (*is*), as in (2a,b), or the main verb (*ate*), as in (1). In a question, on the other hand, the tensed auxiliary verb precedes the subject, as in (2b). In a command, there is neither a subject nor tense; the bare verb begins the sentence, as in (2c). A combination of syntactic (word order) and morphological (presence or absence of tense inflection) differences signals declarative, interrogative, and imperative sentence types. Contrast this with the situation in Lakhota, a Siouan language of North America.[2]

(3) a. Hokšíla ki agúyapi ki yúta-he (yeló). Statement
boy the bread the eat-PROG DECL
'The boy is eating the bread.'
b. Hokšíla ki agúyapi ki yúta-he he? Question
boy the bread the eat-PROG INT
'Is the boy eating the bread?'
c. Agúyapi ki yúta ye! Command
bread the eat IMP
'Eat the bread!'

Lakhota, unlike English, expresses these different types of sentences by simply adding particles at the end of the sentence; no change is made in their syntactic structure, except for the omission of the subject in the command in (3c). The direct object NP and the verb are in the same position in all three examples. The optional particle *yeló* in (3a) signals that the sentence is a declarative utterance, i.e., a statement; it also indicates that the speaker is male. The particle *he* in (3b) signals that the sentence is a question (it is neutral with respect to the sex of the speaker), while the particle *ye* in (3c) indicates that the sentence is a command and that the speaker is female. This way of expressing questions, statements, and commands is much more common across the world's languages than the English pattern in (2), and the contrast between the two illustrates how the same communicative functions can be carried out in very different ways in different languages.

Functionalists normally focus on these linguistic functions from either of two perspectives. They will be referred to as the "pragmatics" perspective and the "discourse" perspective. The first perspective concentrates on the meaning of and the conditions on the appropriate use of different speech acts. The work is based on Searle's (1969, 1985) theory of speech acts and Grice's (1975, 1989) theory of the logic of conversation; Levinson (1983) provides an excellent overview. As an example of the kind of problem which this aspect of functional linguistics addresses, consider the following utterance.

(4) Can you pass the salt?

Taken literally, this question is about the addressee's ability to give the speaker the salt; its literal meaning can be paraphrased as "Are you able to give me the salt?". This is not how it is normally interpreted, however; it is normally understood as a request, not a question, and if the addressee simply answered "yes" without handing the speaker the salt, such a response would be considered impertinent, rude, or smart-alecky. The theories of Grice and Searle make it possible for linguists (and philosophers of language) to explain how a sentence with one form (that

of a question) and a clearcut literal meaning can be interpreted in context as a different kind of speech act with a rather different meaning.

The second perspective is concerned with the construction of discourse and how grammatical and other devices are employed to serve this end. As a simple example of this, consider the problem of keeping track of referents in discourse. When a speaker constructs a text about a number of states of affairs, each of which contains a number of participants, how does he or she code the referents so that the interlocutors can keep them apart but also keep track of the same referents that appear in more than one state of affairs? This problem is illustrated in the following English examples.

(5) a. Mary called Sam, and she talked to him for an hour. He scolded her for refusing to help her sister at the party, and she replied that she had been too busy.

b. Mary called Sam, talked to him for an hour, was scolded by him for refusing to help her sister at the party, and replied that she had been too busy.

The two participants to be tracked are Mary and Sam, and in (5a) they are unambiguously referred to by third-person pronouns that are differentiated in terms of gender. Hence *she* or *her* always refers to Mary and *he* or *him* to Sam.[3] The situation is somewhat different in (5b); there are nouns or pronouns referring to Mary only in the first (*Mary*), third (*her*) and fourth (*she*) clauses, and yet she is clearly a participant in the state of affairs expressed by each clause. In this sentence, Mary is being tracked by syntactic means: the NP *Mary* is the subject of each clause, and this *NP* is omitted after the initial one. In this multiclause construction, a missing subject must be interpreted as being the same as the subject of the first clause in it; hence all of the clauses are construed as having the NP referring to Mary as the subject. The other participant, Sam, is tracked by means of a gender-marked pronoun, just as in (5a). This involves many of the central mechanisms of English clause-internal grammar: the voice of the verb (active vs. passive), grammatical relations (subject vs. nonsubject), and case marking (nominative [*he, she*] vs. accusative [*him, her*]). English thus has two different ways of keeping track of referents in discourse: the gender-marked pronoun system in both (5a) and (5b), and the syntactic system in (5b). Why should it need the system in (5b), when the one in (5a) seems to work just fine? Consider the slightly different examples in (6).

(6) a. Bill called Sam, and he talked to him for an hour. He scolded him for refusing to help his sister at the party, and he replied that he had been too busy.

b. Bill called Sam, talked to him for an hour, was scolded by him for refusing to help his sister at the party, and replied that he had been too busy.

In these examples both participants are male, and therefore the pronouns *he* and *him* are used to refer to both of them. The result in (6a) is serious ambiguity; who, for example, scolded whom? Either Bill or Sam could have done the scolding. In (6b), on the other hand, there is much less ambiguity. The NP referring to Bill must be interpreted as the subject of each of the clauses in the construction, and therefore the pronoun *him* in nonsubject position is interpreted as referring to Sam. The only real ambiguity is with respect to whose sister it is; it could be either Bill's or Sam's. Thus in this case the syntactic referent-tracking mechanism yields less ambiguity than the gender-marked pronoun system. Different languages use different referent-tracking systems: some use gender-marked pronouns primarily, some use syntactic mechanisms primarily, and some use combinations of them (see Van Valin 1987; Comrie 1989, 1994; A. A. Kibrik 1991).

It was mentioned above that many of the basic mechanisms crucial to clause-internal grammar are involved in reference tracking, and this highlights an important aspect of functional analysis. Voice constructions like passive or grammatical relations like subject and direct object are not treated as purely formal grammatical entities; rather, they are analyzed in terms of the functions they serve. With respect to voice constructions, in some languages they are part of a

referent-tracking system, as in English, while in other languages they are not. If a language has a syntactic referent-tracking system, then grammatical relations like subject will be centrally involved in it; in languages which use a gender-marked pronoun system only, then neither voice nor grammatical relations will serve any significant role in referent tracking. In functional linguistic analysis, forms are analyzed with respect to the communicative functions they serve, and functions are investigated with respect to the formal devices that are used to accomplish them. Both forms and functions are analyzed, not just functions. The interplay between form and function in language is very complex and is the prime focus of functional linguistics.

Even though examples from only two languages have been given so far, it should be clear that crosslinguistic comparison is a very significant feature of functional linguistics. As we will see in the next section, it is not a necessary part of this approach, as there are schools of functional linguistics which are not typologically oriented. In the United States in particular, the development of functional linguistics has gone hand in hand with the expansion of the study of language typology and universals. While there are typologists who are not functionalists, the combination of typology and functionalism is not just an accidental pairing of unrelated endeavors. Many of the major figures in the development of functional linguistics in the United States have worked on languages in which the grammatical marking of communicative functions is more obvious and direct than it is in English, the language on the basis of which most theorizing in linguistics in the USA was done up through the end of the 1970s. For instance, since the mid-1950s linguists have recognized that the NP referring to the topic of the discourse (roughly, the participant the discourse is primarily about) is accorded special treatment in the grammatical systems of some languages (see Chao 1955; Hockett 1958; Lambrecht 1994). In the examples in (5) the sentences are about Mary, while in (6) it is Bill who is the topic. Two languages in which the notion of topic plays an important role are Mandarin Chinese and Japanese; in Mandarin, topic NPs may be given special syntactic treatment, and in Japanese they are marked by a special particle, *wa*.

(7) a. Nèi xie shù, shùshēn dà. — Mandarin
those CL tree tree.trunk big — (Li and Thompson 1976)
'Those trees, the trunks are big.'

b. Nihon wa, Tokyo ga sumi-yoi. — Japanese
Japan TOP SUBJ easy-live — (Kuno 1973)
'As for Japan, Tokyo is easy to live in.'

Linguists such as Kuno, Li, Thompson and others took the insights derived from their study of Mandarin, Japanese, and other so-called "exotic" languages and applied them to the analysis of English and other more familiar languages; there they found functional motivations for grammatical phenomena, albeit not always coded as directly as in these languages. Hence the investigation of languages from Asia, Africa, Australia, and the Americas led to insights about the interaction of form and function in language that led directly to the development of functional linguistics in the United States. Functional approaches also arose in Soviet/Russian linguistics based on the study of the non-Slavic languages of the former USSR (A. E. Kibrik 1979, 1985; Bondarko 1991); these were undoubtedly influenced by the well-established Prague-based tradition of Slavic functional linguistics, which will be discussed in the next section.

3 A Brief Look at the Development of Linguistic Theory in the Twentieth Century

At the beginning of the chapter it was noted that many linguists hold that the fact that language is used for communication is largely irrelevant to its analysis. How did such a view arise? The answer lies in the theoretical development of linguistics in the twentieth century. The primary

concern of linguists such as Franz Boas in the USA and Ferdinand de Saussure in Europe at the turn of the century was to lay out the foundations for linguistic science and thereby to define clearly and explicitly the object to be investigated in linguistic inquiry. Culminating in Boas (1911) and Saussure (1917[1959]), this work defined what came to be known as structural linguistics. Saussure drew a fundamental contrast between language (*langue*) and speaking (*parole*): language is a system of signs, whereas speaking is the use of the system on particular occasions. A linguistic sign is the association of a sound (signifier) and a meaning (signified), e.g., the Japanese signifier /inu/ has the signified 'dog,' while the English signifier /əd/ (orthographic *-ed*) has the signified 'past tense.' Saussure argued that the proper subject for linguistic investigation is the system of signs, not the use of the system. Bloomfield (1933) proposed a similar distinction: grammar (the linguistic system) vs. meaning (the use of the system on particular occasions). He, too, argued that linguistic analysis should concern itself only with grammar.

What is the nature of the linguistic system? Saussure proposed that there are two fundamental relations among signs which define a structural system: co-occurrence (syntagmatic) and substitution (paradigmatic). The English sign *-ed*, for example, is in a syntagmatic relation with the verbs that it appears suffixed to, e.g., *load*, *pit*, and *include*, and it is in a paradigmatic relation with the other suffixes that can occur on these verbs, e.g., *-s* 'present tense,' *-en* 'past participle,' or *-ing* 'present participle.' Similarly, Lakhota *agúyapi* 'bread' from (3) is in a syntagmatic relation with *ki* 'the' and a paradigmatic relation with *hokšíla* 'boy,' since both *agúyapi* and *hokšíla* can co-occur with *ki* 'the.' In (3a,b), *hokšíla ki* 'the boy' is syntagmatically related to both *agúyapi ki* 'the bread' and *yúta* 'eat,' and it is paradigmatically related to other Lakhota NPs which can co-occur with these two other elements, e.g., *wičháša ki* 'the man' or *wíyą wą* 'a woman,' as shown in (8).

(8) a. Wičháša ki agúyapi ki yúta-he (yeló).
man the bread the eat-PROG DECL
'The man is eating the bread.'

b. Wíyą wą agúyapi ki yúta-he (yeló).
woman a bread the eat-PROG DECL
'A woman is eating the bread.'

Syntagmatic relations define the frame in which paradigmatic relations exist, and the elements in a paradigmatic relation to each other constitute classes which are in syntagmatic relation to each other. To continue the Lakhota example, "noun + article" constitute a syntagmatic frame, i.e., they co-occur with each other as a regular pattern in the language. Each of the constituents of this pattern, namely "noun" and "article," are themselves names for substitution classes; that is, in terms of the examples we have seen, *wičháša* 'man,' *wíyą* 'woman,' *hokšíla* 'boy,' and *agúyapi* 'bread' can be substituted for each other in the "noun" position in the frame, and *ki* 'the' and *wą* 'a' can be substituted for each other in the "article" position. Syntagmatic (co-occurrence) and paradigmatic (substitution) relations among signs constitute the structure of language, and it is this structure, and not the way signs are used in speaking, that is the proper domain of linguistic study, according to Saussure and Bloomfield.

Chomsky (1965) proposed a distinction analogous but not identical to Saussure's and Bloomfield's, namely competence vs. performance. Competence refers to a native speaker's knowledge of his or her native language, and performance is how a speaker puts that knowledge to use on particular occasions. Performance is very close to Saussure's *parole* and Bloomfield's "meaning," but competence includes not only the linguistic system but also native speakers' knowledge of it. Hence it adds a cognitive dimension to linguistics that had not been emphasized by Saussure and had been explicitly denied by Bloomfield. For Chomsky, the proper domain of linguistic inquiry is competence only.

How do Saussure's, Bloomfield's, and Chomsky's distinctions relate to the issue raised in the introduction, namely, the primary function of language? Since *parole*/meaning/performance

concerns the use of language, and since one of these uses is surely for communication among humans, it is natural to associate the communicative function of language with *parole*/meaning/performance. As we have seen, all three theorists maintain that linguistics is not concerned with the analysis of *parole*/meaning/performance but rather with the study of *langue*/grammar/competence. Hence it is but a short leap to the conclusion that the communicative functions of language are irrelevant to the analysis of language structure (*langue*/grammar/competence). Given that Saussure is generally acknowledged to have laid the foundations for the modern study of language, it is consequently not surprising that many linguists have adopted this view. Thus, a view of language that might seem puzzling to nonlinguists arises rather naturally out of the way linguistic theory has developed in this century.

Does this mean that functional theories are necessarily theories of *parole*/meaning/performance? The answer is, for the most part, "no." Foley and Van Valin (1984) make this point explicitly.

> It must be emphasized that functional theories are *not* performance theories. That is, they seek to describe language in terms of the *types* of speech activities in which language is used as well as the *types* of constructions which are used in speech activities. They do not attempt to predict the actual *tokens* of speech events. In other words, the theories seek to describe the interaction of syntax, semantics and pragmatics in types of speech activities; they do not try to predict the occurrence of particular constructions in actual speech events. They are theories of systems, not of actual behavior. (Foley and Van Valin 1984: 15 [emphasis in original])

How can the various communicative functions of language discussed in Section 2 be incorporated into the study of language structure? The two fundamental relations defining a structural system are co-occurrence (syntagmatic) and substitution (paradigmatic), as mentioned above. The co-occurrence relations among substitution classes constitute a level of structure. If the elements in the substitution classes are phonemes, then the syntagmatic combinations of phonemes are morphemes. If the elements in the substitution classes are morphemes, the syntagmatic combinations are words. If the elements in the substitution classes are words, the syntagmatic combinations are sentences. This is the extent of the study of *langue*/grammar/competence as practiced by Saussure, Bloomfield, and Chomsky. But it is possible to extend the analysis further: if the elements in the substitution classes are sentences, then the syntagmatic combinations are discourses or kinds of speech events. In analyzing sentence types in terms of the kinds of speech events or discourse they can occur in, one is analyzing their communicative function. So, for example, the examples in (2), (3), and (4) all occur in specific types of speech acts. The examples in (5) and (6), on the other hand, involve sentences with particular properties within a discourse context. Hence it is in fact possible to extend the study of *langue*/grammar/competence to take the communicative functions of linguistic forms into account. This is what Foley and Van Valin were getting at in the extract above: it is possible to analyze the potential contexts in which constructions appear, in order to uncover the contextual constraints on their distribution, an idea that goes back to Harris (1952). We can take passive constructions to exemplify this point. In the active voice in English, the doer of the action is the subject, while in the passive voice the NP referring to the participant affected by the action is the subject. In (5b) and (6b) passive is used in the third clause. The subject, which is the topic of the mini-discourse, is not the doer of the action of the verb in that clause. The construction in (5b) and (6b) requires that the topic be the subject of each sentence in it, and therefore passive must be used in the third sentence. This suggests that there is a connection between the topicality of participants and the occurrence of the passive construction, i.e., when the doer of the action is less topical than the other participant, a passive is favored, because it permits the more topical participant to appear as subject. Subjects in English and many other languages are typically topic-like, although there are instances of nontopic subjects. The overwhelming tendency in languages is for the NP referring to the topic to come first in a sentence, followed by elements introducing new information into the context. The following possible question-answer pairs illustrate this.

(9) a. Who did Sally slap?
b. She slapped Pat.
b´. Pat was slapped by Sally/her.

While the sentences in (9b) and (9b´) are perfectly grammatical English sentences, they are not equally good as answers to the question in (9a); (9b) is much better than (9b´). (The most likely answer would be *Pat* by itself, but the whole sentence is included to help illustrate the point.) Part of the reason for this difference in appropriateness derives from discourse factors. The question in (9a) establishes Sally as the topic and also that slapping occurred; the new information requested is the identity of the person slapped. The sentence in (9b) presents the elements expressing established information first followed by the NP *Pat*, which is the answer to the question. The sentence in (9b´), on the other hand, presents the new information first followed by the established information, with the topic NP last in the sentence. Hence it is inappropriate in the context created by the question in (9a). This account is somewhat oversimplified, but it nevertheless illustrates how linguistic analysis can be extended to take communicative functions into account.

The idea of extending linguistic analysis to include communicative functions was first proposed by Czech linguists. Virtually all contemporary functional approaches trace their roots back to the work of the Czech linguist Mathesius in the 1920s as part of the Prague School (Mathesius 1928, 1929). He and his successors developed the theory of functional sentence perspective. They were the first to fully develop the observation that the elements expressing more established information (what was earlier called the "topic," what the Pragueans call the "theme") precede the elements expressing new information (what is often called the "focus" and what Pragueans call the "rheme"). This a salient feature of Slavic languages, as the following examples from Russian (Comrie 1979) show. In the translations, the focus is in small caps, and the square brackets group the topical elements and focal elements together.

(10) a.	Q:	[Kto]	[zaščiščajet Viktor-a]?	'Who defends Victor?'
		who.NOM	defends Victor-ACC	
		FOCUS	TOPIC	
		A:[Viktora	zaščiščajet] [Maksim-Ø].	'MAXIM defends Victor.'
		Victor- ACC	defends Maxim-NOM	
		TOPIC	FOCUS	
b.	Q:	[Kogo]	[zaščiščajet Maksim-Ø]?	'Who(m) does Maxim defend?'
		who. ACC	defends Maxim-NOM	
		FOCUS	TOPIC	
	A:	[Maksim-Ø	zaščiščajet] [Viktor-a].	'Maxim defends VICTOR.'
		Maxim-NOM	defends Victor-ACC	
		TOPIC	FOCUS	
c.	Sp1:	[Maksim-Ø]	[ubivajet Aleksej-a].	'Maxim KILLS ALEXEI.'
		Maxim-NOM	kills Alexei-ACC	
		TOPIC	FOCUS	
	Sp2:	[A Viktor-a]?		'And VICTOR? [i.e., 'What is happening to Victor?'"]
		and Victor-ACC		
		FOCUS		
	Sp1:	[Viktor-a	Maksim-Ø] [zaščiščajet].	'Maxim DEFENDS Victor.'
		Victor-ACC	Maxim-NOM defends	
		TOPIC	FOCUS	

Aside from the fact that question words like *kto* and *kogo* 'who' occur at the beginning of the sentence, as they do in many languages, the ordering of elements is topic (theme) followed by focus (rheme). It was noted in Section 2 that the study of so-called "exotic" languages by English-speaking linguists had led to insights about the functional motivation for grammatical phenomena, but here the crucial insight derives from the native language of the linguists. The theory of functional sentence perspective was developed primarily with respect to the analysis of Slavic languages, but its ideas have been applied by other linguists to a range of phenomena in many languages.[4] This theory was first brought to the attention of English-speaking linguists in Halliday (1967); Kuno (1972a, 1972b) and Chafe (1972) applied them to issues that were of concern to theoretical linguists in the USA at that time. By the end of the 1970s, a number of functional approaches were emerging in both the USA and Western Europe.

4 Functional Approaches

There is a great diversity of views among those who label themselves as functionalists. One of the curious features of functionalism in linguistics is the apparent paucity of explicitly articulated, named theories. There are really just three: Functional Grammar [FG] (Dik 1978, 1989) and its newer version, Functional Discourse Grammar (Hengeveld and Mackenzie 2008); Systemic Functional Grammar [SFG] (Halliday 1967, 1994); and Role and Reference Grammar [RRG] (Foley and Van Valin 1984; Van Valin 1993, 2005; Van Valin and LaPolla 1997). Nichols (1984) presents a survey of functionalist approaches which usefully categorizes them as extreme, moderate, and conservative. Her descriptions of each are still valid and are given below.

> The conservative type merely acknowledges the inadequacy of strict formalism or structuralism, without proposing a new analysis of structure… The moderate type not only points out the inadequacy of a formalist or structuralist analysis, but goes on to propose a functionalist analysis of structure and hence to replace or change inherited formal or structural accounts of structure… Extreme functionalism denies, in one way or another, the reality of structure qua structure. It may claim that rules are based entirely on function and hence there are no purely syntactic constraints; that structure is only coded function, or the like. (Nichols 1984: 102–3)

Conservative functionalism, as exemplified in the work of Kuno (e.g., 1972a, 1972b, 1975, 1987) and Prince (e.g., 1981a, 1981b), seeks to augment standard formal analyses with functional principles, thereby creating an additional functional "component" or "module" in the grammar. Kuno (1987) is very explicit on this point.

> Functional syntax is, in principle, independent of various past and current models of grammar… Each theory of grammar must have a place or places where various functional constraints on the well-formedness of sentences or sequences of sentences can be stated, and each can benefit from utilizing a functional perspective in the analysis of concrete syntactic phenomena. Therefore, in theory there is no conflict between functional syntax and say, the government and binding theory of generative grammar. (Kuno 1987: 1)

These approaches assume the same basic notion of grammatical structure that formal theories do and propose constraints or rules that either supplement or in some case even replace purely structure-oriented rules. They do not challenge the fundamental assumptions of formal theories, and therefore they represent an extension of them rather than an alternative to them. Indeed, out of this work has evolved a number of formal theories of information structure (see Section 5 below).

Moderate functional theories do reject the assumptions of formal theories such as Chomsky's and are presented as alternatives to them. Two of the theories mentioned above, FG and RRG, are

moderate functional theories. These theories reject the conceptions of grammatical structure that underlie formal theories, but each proposes a different replacement view of structure. However, they do not deny the validity of the notion of structure *per se* and do not claim that all grammatical structure is reducible to discourse structure or some other functional notion(s). Rather, they view grammatical structure as strongly influenced by semantics and pragmatics and undertake to explore the interaction of structure and function in language. The following are characterizations of moderate functionalist views of language. First, Dik (1991) characterizes the FG view of language as follows.

> [A] language is considered in the first place as an instrument for communicative verbal interaction, and the basic assumption is that the various properties of natural languages should, wherever this is possible, be understood and explained in terms of the conditions imposed by their usage. The language system, therefore, is not considered as an autonomous set of rules and principles, the uses of which can only be considered in a secondary phase; rather it is assumed that the rules and principles composing the language system can only be adequately understood when they are analyzed in terms of conditions of use. In this sense the study of language use (pragmatics) precedes the study of the formal and semantic properties of linguistic expressions. (Dik 1991: 247)

Second, Van Valin (1993) lays out the basic assumptions of RRG as follows.

> RRG takes language to be a system of communicative social action, and accordingly, analyzing the communicative functions of grammatical structures plays a vital role in grammatical description and theory from this perspective… Language is a system, and grammar is a system in the traditional structuralist sense; what distinguishes the RRG conception…is the conviction that grammatical structure can only be understood with reference to its semantic and communicative functions. Syntax is not autonomous. In terms of the abstract paradigmatic and syntagmatic relations that define a structural system, RRG is concerned not only with relations of co-occurrence and combination in strictly formal terms but also with semantic and pragmatic co-occurrence and combinatory relations. (Van Valin 1993: 2)

The rules and constraints proposed in FG and RRG bear little resemblance to those proposed in generative theories, and therefore these theories do not complement formal theories but rather are alternatives to them. Both of these theories are strongly typologically oriented. RRG, for example, grew out attempts to answer the following questions: (i) what would linguistic theory look like if it were based on the analysis of Lakhota, Tagalog, and Dyirbal, rather than on the analysis of English?, and (ii) how can the interaction of syntax, semantics and pragmatics in different grammatical systems best be captured and explained? (Van Valin 2005: 1).[5]

Extreme functionalism, as manifested in the works of Hopper (1987), Bybee and Hopper (2001), Thompson (2002), and others, rejects the validity of any notion of structure other than that of discourse structure and seeks a radical reduction of grammar to discourse. On this view, grammar is strongly motivated by discourse, and the emphasis on the primacy of discourse leads even to the rejection of semantics as a valid part of linguistic investigations, where "semantics" is understood as the study of the meaning of forms independent of their discourse function(s). It is not clear whether the result of this kind of analysis yields something that could be called "grammar" in any accepted sense. Extreme functionalism abandons the basic Saussurean conception of language as a structural system, which underlies structural and generative linguistics, as well as conservative and moderate functionalism.

An alternative label for this anti-Saussurean view of language is "usage-based," a term originally proposed by Langacker (1987), one of the founders of cognitive linguistics and a leading figure in it. Usage-based models specifically deny the Saussurean distinction between *langue* and *parole*, or more relevantly, between grammar and use, claiming that grammar arises out of and can only be explained in terms of language use (e.g., Bybee 2006; Barlow and Kemmer 2000;

Tomasello 2003). Langacker and Tomasello are both cognitive linguists, and this signifies that the range of linguists who identify themselves as working in a usage-based framework includes both cognitivists and functionalists. This approach can be seen as uniting cognitivism and functionalism in certain respects.

Falling somewhere between moderate and extreme functionalism is SFG, which takes a strongly discourse-oriented view of language, but which nevertheless does not deny either the reality of structure in language nor the Saussurean foundations of modern linguistics. SFG is a "top-down" analytic model which starts with discourse and works "down" to lower levels of grammatical structure. Halliday (1985) maintains that the ultimate explanations for linguistic phenomena are to be found in language use.

> Language has evolved to satisfy human needs; and the way it is organized is functional with respect to these needs – it is not arbitrary. A functional grammar is essentially a 'natural' grammar, in the sense that everything in it can be explained, ultimately, by reference to how language is used. (1985: xiii) … The orientation is to language as a social rather than an individual phenomenon, and the origin and development of the theory have aligned it with the sociological rather than psychological modes of explanation. At the same time it has been used within a general cognitive framework… (Halliday 1985: xxx)

SFG is less concerned with issues of sentence grammar than FG and RRG, and more with discourse structure. Hence it falls toward the more extreme end of the spectrum, and it could be seen as a pioneer of usage-based linguistics. Butler (2003) presents an in-depth presentation and critical comparison of these three theories.

Among the three approaches, it could be argued that conceptually the biggest gap is between extreme functionalism and all the others, since it represents the greatest departure from the mainstream currents of twentieth-century linguistics. There is a context in which extreme and moderate functionalism fall together in opposition to conservative functionalism, however. This is the issue of the relationship between the functionalist and generative theoretical agendas. As the quote from Kuno earlier in this section makes clear, there is no inherent conflict between the goals of generative grammar and those of conservative functional syntax, but this is not the case with moderate and extreme functionalism. The extreme view rejects the generative enterprise and the questions it deals with altogether; for its adherents, the issues raised by generative researchers are pseudo-problems created by an invalid methodological approach to language. Moderate functionalists have a rather different perspective. Their agenda is broader than that of generative linguistics, since it is not limited to issues of sentence grammar but also includes discourse and other pragmatic issues, and therefore the moderate functionalist agenda *subsumes* the formalist agenda at the same time that it transforms it in terms of functional categories and relations. Thus in RRG, for example, research has focused not only on discourse-related issues like reference tracking but also on formalist issues like constraints on *wh*-question formation and relative clause formation (Van Valin 1995; Van Valin and LaPolla 1997; Van Valin 2005).

One of the most salient features of Chomskyan (but not all formal) linguistics is the goal of describing a native speaker's grammatical competence and explaining the acquisition of language by children. Not surprisingly, functionalist approaches vary with respect to their stand on these issues. Since conservative functionalists basically follow general generativist doctrine, they too subscribe to this goal, and they follow the standard Chomskyan view regarding the existence of an autonomous language faculty. Moderate functionalist theories all adopt this goal, with the reinterpretation of grammatical competence as communicative or textual competence, and at least some (RRG) expressly reject the Chomskyan autonomy hypothesis. Van Valin (1991, 1994, 1998, 2001) and Van Valin and LaPolla (1997) present a model of the acquisition of syntax which does not assume an autonomous language acquisition device. Opinion among extreme functionalists varies on this issue; Hopper (1987) explicitly denies the validity of any psychological

interpretation in linguistics, while Bates and MacWhinney (1982, 1987, 1989) develop what they call the "competition model" to account for language acquisition, and Tomasello (2003) presents a usage-based account of language acquisition.

5 Formal vs. Functional Approaches to Language

This discussion began with contrasting views on the primary function of language, and a dichotomy was set up between those linguists who believe it to be communication and take the communicative functions of language to be important for its analysis (the functionalists) and those who at the very least consider the communicative functions of language to be irrelevant to its analysis, following Chomsky (the formalists). This contrast, as one might suspect, is rather oversimplified, and when one scans the topics that formalists and functionalists investigate, the distinctions become somewhat blurred. Until 20 years ago only functionalists talked about referent tracking, discourse, and information structure (topic, focus), among other issues, but that has changed significantly. There are formal theories of discourse and information structure, e.g., Kamp and Reyle (1993), Vallduví (1992), and Erteschik-Shir (1997), and analyses of the role of notions like topic in the syntax of different languages, e.g., É. Kiss (1987, 1994). The approaches of Vallduví, Erteschik-Shir and É. Kiss can be viewed as the descendants of the conservative functionalists discussed earlier. Until the early 1980s the problems of so-called "exotic" languages were primarily the province of typologists and functionalists, but since then linguists of all theoretical persuasions have begun to investigate them. Whereas 20 years or so ago it was possible to identify a formalist or a functionalist merely by the problems they investigated, this is no longer the case today.

What, then, distinguishes formalists from functionalists? There is one fundamental difference which sets functionalists of all persuasions off from formalists, and there is a second distinction which separates extreme functionalists from both formalists and conservative and moderate functionalists. The first difference concerns the type of explanatory criteria that the approach recognizes. The following table, from Van Valin and LaPolla (1997), summarizes the relevant types of explanatory criteria.

Table 8.1 Types of Explanatory Criteria

Domain to be Explained	**Theory-Internal Criteria**	**External Criteria**	
		Language-Internal	*Language-External*
SYNTAX	Economy Motivation Predictiveness	Phonology Semantics Pragmatics Processing	Reasoning Categorization Perception …

The label "theory-internal" refers to the fact that within a particular domain, e.g., syntax, the criteria are applied to competing analyses within that domain; it does not mean that they are specific to any particular linguistic theory. They are explicated briefly in (11).

(11) Theory-internal explanatory criteria
   a. Economy (also known as "Occam's Razor"): Is it the simplest account?
   b. Motivation: Are the main explanatory constructs independently motivated or are they specific to the problem at hand?
   c. Predictiveness: Do the hypotheses predict phenomena beyond those for which they were formulated?

If an approach restricts itself to theory-internal criteria only, then syntactic phenomena are explained in syntactic terms, semantic phenomena in semantic terms, phonological phenomena in phonological terms, etc. This is often expressed in terms of the thesis of the autonomy of syntax, and it applies to semantics and phonology as well; phenomena in each domain are to be explained in terms of constructs, rules or principles which involve elements in that domain only. The external explanatory criteria involve factors outside of the domain being studied, and they can be internal or external to language itself. Invoking phonetics to account for some phonological phenomenon is an example of permitting language-internal external criteria in explanation, whereas invoking some feature of the human perceptual system to account for some phonological phenomenon would be an instance of using language-external external criteria in explanation.

In a detailed discussion of the contrast between formal and functional linguistics, Newmeyer (1998) argues this difference is more apparent than real and that there is no principled reason why external explanatory criteria could not play a role in formal theories; rather, it is the case that formal linguists have not concerned themselves with external criteria. Contra Newmeyer, this contrast does follow from fundamental differences between the general approaches: Chomskyan formal linguistics has long maintained that syntax was an autonomous component of language, and language is an autonomous component of the mind/brain; hence these claims about autonomy make appealing to external explanatory criteria very problematic. Newmeyer (2003) explicitly defends the Saussurean position against the usage-based view, arguing that grammar and usage are distinct.

As is evident from the citations earlier from Kuno, Dik, Van Valin, and Halliday, functional approaches look to semantics and pragmatics as the basis for explanations of syntactic phenomena. Formal approaches, on the other hand, restrict themselves to theory-internal criteria in explanation, for the most part. When formal and functional accounts of the same phenomena are compared, this contrast stands out clearly. For example, the explanation for the difference in grammaticality in (12) is quite distinct in the two approaches. These sentences involve sentence-internal pronominalization, and the issue is whether a particular lexical noun and a particular pronoun can be interpreted as coreferential (identical subscripts indicate intended coreference).

(12) a. As for his$_i$ sister, Tom$_i$ hasn't talked to her in three weeks.
a´. *As for his$_i$ sister, she hasn't talked to Tom$_i$ in three weeks.
b. It is his$_i$ sister that Tom$_i$ hasn't talked to in three weeks.
b´. It is Tom's$_i$ sister that he$_i$ hasn't talked to in three weeks.

The fact to be explained here is why coreference between *his* and *Tom* is very difficult, if not impossible, to get in (12a´) but possible in the other sentences. In standard formal accounts, e.g., Chomsky (1981), the explanation is stated in terms of the relative positions of the lexical noun and the pronoun in the syntactic phrase-structure tree representing the structure of the sentence. In functional accounts, e.g., Kuno (1975) and Bolinger (1979), the difference is attributed to the different information structures in the sentences, i.e., differences in which NP functions as topic and which as focus. Chomsky defines pronominalization as a syntactic phenomenon, and therefore only syntactic factors are relevant to its explanation; when competing syntactic accounts of pronominalization are evaluated, only the theory-internal criteria are employed. For functionalists like Kuno and Bolinger, on the other hand, semantics and pragmatics can be brought into the explanation, and competing accounts would be evaluated with both theory-internal and theory-external criteria. The centrality of external explanations for linguistic phenomena is a point that all functionalists agree on.

The second distinction which distinguishes extreme functionalists from the rest concerns the role of theory in linguistics. Virtually all formal linguists are strongly committed to working

within a well-defined theoretical framework, but this is not the case with functionalists. Conservative functionalists, who view their work as augmenting formal grammars, fall in with formalists on this point. Many moderate functionalists are likewise theoretically oriented, as witnessed by the development of theories like RRG and FG/FDG, which employ technical metalanguages and explicit representations of the relevant syntactic, semantic, and pragmatic phenomena they investigate. Extreme functionalists, on the other hand, deny the validity of functional theories and maintain that true functional theories are impossible. Givón (1989) argued that all theories are inherently formal, and therefore that a functional theory was a contradiction in terms. They also view the use of any kind of explicit notations or representations as inherently formalist and reject them as well. From an extreme functionalist perspective, RRG and FG (and presumably FDG) do not even count as functional approaches, because of their commitment to theory development and use of explicit notation and representations.

6 Conclusion

The label "functional linguistics" is a cover term for a complex web of ideas and methodologies, many of which are more distant from each other than they are from many formalist ideas. Work by conservative functionalists has yielded important insights regarding the pragmatic nature of many syntactic constraints, but they do not address the crucial question of the nature of structure in language, particularly syntactic structure, since they assume a generative account of structure. Indeed, conservative functionalist ideas have evolved into widely accepted components of contemporary mainstream formal theories. Extreme functionalists have uncovered many important generalizations about discourse structure, information flow, and the discourse functions of grammatical forms, but by rejecting the notion of language as a structural system they, and some of the current proponents of usage-based approaches, have, like the conservative functionalists, avoided one of the central questions of linguistic theory, that of the nature of linguistic structure.[6] Only moderate functionalists have attempted the difficult task of proposing alternative conceptions of linguistic structure and developing explanatory theories. While there has been some convergence between the work of conservative and moderate functionalists on the one hand, and many formalists on the other, they are nevertheless distinguished by their respective views on what counts as an explanation. All functionalists agree that language is a system of forms for conveying meaning in communication and therefore that in order to understand it, it is necessary to investigate the interaction of structure, meaning, and communicative function.

SUGGESTIONS FOR FURTHER READING

Darnell, Michael, Moravcsik, Edith, Newmeyer, Frederick, and Wheatley, Kathleen (eds) (1999). *Functionalism and Formalism in Linguistics* (2 vols.). Amsterdam/Philadelphia, PA: John Benjamins.

Givón, T. (1995). *Functionalism and Grammar*. Amsterdam/Philadelphia, PA: John Benjamins.

RELEVANT JOURNALS

Language
Studies in Language
Functions of Language

EMERGING TRENDS AND RESEARCH QUESTIONS

1. What role do the phenomena investigated by functionalists play in language production and comprehension? Can they be incorporated into models of sentence processing?
2. Can functionalist approaches provide motivation for concepts that are central to formal analyses?

NOTES

1. It is often asserted by advocates of Chomsky's view that science leads to results that defy common sense, the prime example being modern physics. However, the counterintuitive results of special relativity and quantum mechanics deal with phenomena outside the range of human experience, i.e., subatomic particles or objects moving at close to the speed of light. Linguistics does not deal with such phenomena; rather, it deals with what has long been considered the quintessential human phenomenon. Hence it is reasonable to question the denial of the relevance or the importance of the most obvious feature of the phenomenon to be described and explained.
2. Abbreviations used in glosses: CL 'classifier,' DECL 'declarative,' IMP 'imperative,' INT 'interrogative,' PROG 'progressive,' Sp 'Speaker,' SUBJ 'subject,' TOP 'topic.'
3. Strictly speaking, the pronouns could refer to other individuals of the appropriate gender. However, in this and all subsequent examples, we will limit our discussion to the universe of discourse defined by the first sentence in the construction.
4. For an overview of the Prague School, see Vachek (1964, 1983), and for more recent work by Prague School functional linguists, see Sgall, Hajičová, and Panevová (1986) and Firbas (1964, 1992).
5. Tagalog, an Austronesian language, is the national language of the Philippines, and Dyirbal is an Australian Aboriginal language.
6. This is not true of the adherents of Construction Grammar (e.g., Goldberg 2006), who consider this approach to fall under the rubric of usage-based models.

REFERENCES

Barlow, M. and Kemmer, S. (eds) (2000). *Usage-based Models of Language*. Stanford, CA: CSLI.

Bates, E. and MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner and L. Gleitman (eds), *Language Acquisition: The State of the Art* (pp. 173–218). Cambridge: Cambridge University Press.

Bates, E. and MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (ed.), *Mechanisms of Language Acquisition* (pp. 157– 94). Hillsdale, NJ: Lawrence Erlbaum.

Bates, E. and MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney and E. Bates (eds), *The Crosslinguistic Study of Sentence Processing* (pp. 3–76). Cambridge: Cambridge University Press.

Bloomfield, L. (1933). *Language*. New York: Holt.

Boas, F. (1911). Introduction. In F. Boas (ed.), *Handbook of American Indian Languages, vol. I*. Washington: Bureau of American Ethnology.

Bolinger, D. (1979). Pronouns in discourse. In T. Givón (ed.), *Syntax and Semantics, vol. 10: Discourse and Syntax* (pp. 289–310). New York: Academic Press.

Bondarko, A. V. (1991). *Functional Grammar: A Field Approach*. Amsterdam/Philadelphia, PA: John Benjamins.

Butler, C. (2003). *Structure and Function: A Guide to Three Major Structural-Functional Theories* (2 vols.). Amsterdam/Philadelphia, PA: John Benjamins.

Bybee, J. (2006). From usage to grammar: the mind's response to repetition. *Language*, 82, 711–33.

Bybee, J. and Hopper, P. (eds) (2001). *Frequency and the Emergence of Grammar*. Amsterdam/Philadelphia, PA: John Benjamins.

Chafe, W. (1972). *Meaning and the Structure of Language*. Chicago, IL: University of Chicago Press.

Chao, Y. R. (1955). Notes on Chinese grammar and logic. *Philosophy East and West*, 5, 13–31.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*: Cambridge, MA: MIT Press.
Chomsky, N. (1975). *Reflections on Language*. New York: Pantheon.
Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.
Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Reidel.
Comrie, B. (1979). Russian. In T. Shopen (ed.), *Languages and their Status* (pp. 91–150). Cambridge, MA: Winthrop.
Comrie, B. (1989). Some general properties of reference-tracking systems. In D. Arnold, M. Atkinson, J. Durand, C. Grover, and L. Sadler (eds), *Essays on Grammatical Theory and Universal Grammar* (pp. 37–51). Oxford: Clarendon.
Comrie, B. (1994). Towards a typology of reference-tracking devices. Paper presented to the International Symposium on Language Typology, University of Tsukuba, January 1994.
Dik, S. C. (1978). *Functional Grammar*. Amsterdam: North-Holland.
Dik, S. C. (1989). *The Theory of Functional Grammar*, part 1. Dordrecht: Foris.
Dik, S. C. (1991). Functional Grammar. In F. Droste and J. Joseph (eds), *Linguistic Theory and Grammatical Description* (pp. 247–74). Amsterdam/Philadelphia, PA: John Benjamins.
Erteschik-Shir, N. (2007). *Information Structure: The Syntax-Discourse Interface*. Oxford: Oxford University Press.
Firbas, J. (1964). On defining the theme in Functional Sentence Perspective. *Travaux linguistiques de Prague*, 1, 267–80.
Firbas, J. (1992). *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.
Foley, W. A. and Van Valin, R. D., Jr. (1984). *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
Givón, T. (1989). *Mind, Code and Context: Essays in Pragmatics*. Hillsdale, NJ: Lawrence Erlbaum.
Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (eds), *Syntax and Semantics, vol. 3: Speech Acts* (pp. 41–58). New York: Academic Press.
Grice, H. P. (1989). *Studies in the Ways of Words*. Cambridge, MA: Harvard University Press.
Halliday, M. A. K. (1967). Notes on transitivity and theme in English. *Journal of Linguistics*, 3, 37–81 (Pt. I), 199–244 (Pt. II).
Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. Baltimore, MD: University Park Press.
Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*, 2nd edn. London: Edward Arnold.
Harris, Z. (1952). Discourse analysis. *Language*, 28, 1–30.
Hengeveld, K. and Mackenzie, J. L. (2008). *Functional Discourse Grammar: A Typologically-based Theory of Language Structure*. Oxford: Oxford University Press.
Hockett, C. (1958). *A Course in Modern Linguistics*. New York: Macmillan.
Hopper, P. (1987). Emergent Grammar. *Berkeley Linguistics Society Proceedings*, 13, 139–57. Berkeley, CA: University of California.
Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. Hingham, MA: Kluwer.
Kibrik, A. A. (1991). Maintenance of reference in sentence and discourse. In W. P. Lehmann and H. Jakusz Hewitt (eds), *Language Typology 1988* (pp. 57–84). Amsterdam/ Philadelphia, PA: John Benjamins.
Kibrik, A. E. (1979). Canonical ergativity and Daghestan languages. In F. Plank (ed.), Ergativity (pp. 61–78). London: Academic Press.
Kibrik, A. E. (1985). Toward a typology of ergativity. In J. Nichols and A. Woodbury (eds), *Grammar Inside and Outside the Clause* (pp. 268–323). Cambridge: Cambridge University Press.
É. Kiss, K. (1987). *Configurationality in Hungarian*. Dordrecht: Reidel.
É. Kiss, K. (ed.) (1994). *Discourse Configurational Languages*. Oxford: Oxford University Press.
Kuno, S. (1972a). Functional Sentence Perspective: A case study from Japanese and English. *Linguistic Inquiry*, 3, 269–320.
Kuno, S. (1972b). Pronominalization, reflexivization, and direct discourse. *Linguistic Inquiry*, 3, 161–96.
Kuno, S. (1973). *The Structure of the Japanese Language*. Cambridge, MA: MIT Press.
Kuno, S. (1975). Three perspectives in the functional approach to syntax. *CLS Parasession on Functionalism*, 276–336. Chicago, IL: University of Chicago.
Kuno, S. (1987). *Functional Syntax: Anaphora, Discourse, and Empathy*. Chicago, IL: University of Chicago Press.
Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
Langacker, R. (1987). *Foundations of Cognitive Grammar, vol. 1*. Stanford, CA: Stanford University Press.
Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
Li, C. N. and Thompson, S. A. (1976). Subject and topic: a new typology of language. In C. N. Li (ed.), *Subject and Topic* (pp. 459–89). New York: Academic Press.

Mathesius, V. (1928). On linguistic characterology with illustrations from modern English. *Actes du Premier Congès international de linguistes à la Haye*, 56–63. [Reprinted in J. Vachek (1964), ed., 59–67.]

Mathesius, V. (1929). Functional linguistics. In J. Vachek (ed.) (1983) *Praguiana* (pp. 121–42). Amsterdam/Philadelphia, PA: John Benjamins.

Newmeyer, F. Jr. (1998). *Language Form and Language Function*. Cambridge, MA: MIT Press.

Newmeyer, F. Jr. (2003). Grammar is grammar and usage is usage. *Language*, 79, 682–707.

Nichols, J. (1984). Functional theories of grammar. *Annual Review of Anthropology*, 13, 97–117.

Prince, E. (1981a). Topicalization, Focus-movement, and Yiddish-movement: a pragmatic differentiation. *BLS*, 7, 249–64.

Prince, E. (1981b). Toward a taxonomy of given-new information. In P. Cole (ed.), *Radical Pragmatics* (pp. 223–55). New York: Academic Press.

Saussure, F. de (1917[1959]). *Course in General Linguistics*. New York: McGraw-Hill.

Searle, J. R. (1969). *Speech Acts*. Cambridge: Cambridge University Press.

Searle, J. R. (1985). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.

Sgall, P., Hajičová, E., and Panevová, J. (1986). The Meaning of the Sentence in its Semantic and Pragmatic Aspects, ed. by J. L. Mey. Dordrecht/Boston: D. Reidel.

Silverstein, M. (1976). Shifters, linguistic categories and cultural descriptions. In K. Basso and H. Shelby (eds), *Meaning in Anthropology* (pp. 11–56). Albuquerque, NM: University of New Mexico Press.

Silverstein, M. (1977). Cultural prerequisites to grammatical analysis. In M. Saville-Troike (ed.), *Linguistics and Anthropology* (pp. 139–51). Washington: Georgetown University Press.

Silverstein, M. (1987). The three faces of 'function': preliminaries to a psychology of language. In M. Hickman (ed.), *Social and Functional Approaches to Language and Thought* (pp. 17–38). Orlando, FL: Academic Press.

Thompson, S. (2002). 'Object complements' and conversation: towards a realistic account. *Studies in Language*, 26, 125–64.

Tomasello, M. (2003). *Constructing a Language*. Cambridge, MA: Harvard University Press.

Vachek, J. (ed.) (1964). *A Prague School Reader in Linguistics*. Bloomington, IN/ London: Indiana University Press.

Vachek, J. (ed.) (1983). *Praguiana*. Amsterdam/ Philadelphia, PA: John Benjamins.

Vallduví, E. (1992). *The Informational Component*. New York: Garland.

Van Valin, R. D., Jr. (1987). Aspects of the interaction of syntax and pragmatics: discourse coreference mechanisms and the typology of grammatical systems. In J. Verschueren and M. Bertuccelli-Papi (eds), *The Pragmatic Perspective: Selected Papers from the 1985 International Pragmatics Conference* (pp. 513–31). Amsterdam/Philadelphia, PA: John Benjamins.

Van Valin, R. D., Jr. (1991). Functionalist linguistic theory and language acquisition. *First Language*, 11, 7–40.

Van Valin, R. D., Jr. (1993). A synopsis of Role and Reference Grammar. In R. D. Van Valin, Jr. (ed.), *Advances in Role and Reference Grammar* (pp. 1–164). Amsterdam/Philadelphia, PA: John Benjamins.

Van Valin, R. D., Jr. (1994). Extraction restrictions, competing theories and the argument from the poverty of the stimulus. In S. Lima, R. L. Corrigan, and G. K. Iverson (eds), *The Reality of Linguistic Rules* (pp. 243–59). Amsterdam/Philadelphia, PA: John Benjamins.

Van Valin, R. D., Jr. (1995). Toward a functionalist account of so-called extraction constraints. In B. Devriendt, L. Goosens, and J. von der Auwera (eds), *Complex Structures: A Functionalist Perspective* (pp. 29–60). Berlin: Mouton de Gruyter.

Van Valin, R. D., Jr. (1998). The acquisition of WH-questions and the mechanisms of language acquisition. In M. Tomasello (ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure* (pp. 221–49). Hillsdale, NJ: Lawrence Erlbaum.

Van Valin, R. D., Jr. (2001). The acquisition of complex sentences: a case study in the role of theory in the study of language development. *CLS*, 36-2, 511–31.

Van Valin, R. D., Jr. (2005). *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Van Valin, R. D., Jr. and LaPolla, R. J. (1997). *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Part III Core Fields

9 Linguistic Phonetics

The Sounds of Languages

JOHN LAVER

1 Introduction

Linguistics and phonetics are often characterized as *the linguistic sciences*, implying both similarities and differences. Both linguistics and phonetics are grounded in a basic interest in the nature of human communication. If the subject of *linguistics* is the scientific study of the nature, use, and variety of all aspects of language, the subject of *phonetics* is the scientific study of the nature, use, and variety of all aspects of speech. These are broad definitions of both subjects, and not all linguists and phoneticians would accept such a breadth of scope. But the two subjects have developed so widely in the second half of the twentieth century that a broad view probably better represents the modern diversity of both subjects. What is less controversial is that linguistics and phonetics share a common if partial domain in *phonology*, the study of communicative aspects of spoken language. (In the text below, the first significant mention of a technical term is printed in italics.)

The intersection of linguistics and phonetics in the study of spoken language is visible in the perspectives that each borrows from the other for phonological purposes. Linguistics contributes to phonetics its phonological understanding of the distinctive patterns that make up the coded, conventional aspects of speech which differentiate individual words and other units of spoken language. Phonetics contributes to linguistics its phonetic understanding of the production and perception of the detailed artefacts of speech that embody those significant phonological patterns. Each contribution is complemented by the other. To study formal patterns alone risks becoming over abstract, and losing touch with the physical realities of spoken language. To study the artefacts of speech without due regard for their identity as conventionally coded signals risks losing sight of the communicative motive of spoken language. The name usually given to the study of spoken language from a phonetic perspective, following the example of Ladefoged (1971, 1997), is *linguistic phonetics*.

2 Linguistic Phonetics and General Phonetic Theory

The objective of linguistic phonetics, which most phoneticians would regard as the center of their professional domain, is to describe the phonetic correlates of phonological units of spoken

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

language and their interactions. Another way of putting this is to say that the ultimate task of linguistic phonetics is to give a comprehensive account of speech patterns and their pronunciations in all languages and dialects of the world. To achieve this task, linguistic phonetics draws on *general phonetic theory*, which is the foundation for the phonetician's understanding of how speech is produced, transmitted, and perceived, against a background of a *general phonological theory* of spoken language. The aim of this chapter is to give a compact account of the shape and content of a model of linguistic phonetics within this framework of a general phonetic theory. Within the current volume, Chapter 10 presents a summary view of the shape and content of phonological theory, and the reader is referred to that chapter for definitions of basic phonological concepts used here, such as "phoneme," "allophone," "phonological feature," and "phonological syllable."

More extensive presentations of linguistic phonetics than is possible here are available in Abercrombie (1967), Catford (1977, 1988, 1994), Clark and Yallop (1995), Ladefoged (1993, 1997), Ladefoged and Maddieson (1996), and Laver (1994a). Hardcastle and Laver (1997) offer a comprehensive account of the phonetic sciences, including both linguistic and nonlinguistic aspects. A branch of phonetics with particular relevance to both linguistic phonetics and phonology is *acoustic phonetics*. Recommended publications in acoustic phonetics for readers interested in linguistic communication are Kent and Read (1992), Ladefoged (1971, 1993), and Stevens (1998). Laver (1994b) surveys nonlinguistic interests in phonetics, including *paralinguistic* interests in communication of attitudinal and emotional information through tone of voice, and *extralinguistic* interests in matters such as speaker-characterization. Coulmas (1992) provides a comprehensive account of phonetic and linguistic variation in different *sociolinguistic* speech communities. Goldsmith (1995) gives a wide-ranging review of many different approaches to phonological theory.

3 The Scope of Linguistic Phonetics

A comprehensive approach to linguistic phonetics might entail addressing at least four complementary objectives:

1 describing the phonetic basis for differentiating all contrastive (*phonemic*) and contextual (*allophonic*) patterns in speech which signal the identities of linguistic units in any given language;
2 describing the phonetic regularities which distinguish the speech styles of a given sociolinguistic community from those of others within any given language;
3 describing the idiosyncratic but characteristic phonetic events which distinguish the speech of one member of any given sociolinguistic community from that of other members;
4 describing all recurrent aspects of speech that make one language sound different from others.

All four objectives could be thought relevant to capturing the full extent of the behavioral substance of spoken linguistic communication. Most linguistic phonetic accounts of languages, however, have almost entirely restricted themselves to the first objective. Research by a number of other specialisms has used this first objective as a foundation for pursuing one or more of the other objectives. Sociolinguists interested in the way that speech acts as an index of membership of different communities have investigated the second objective, usually in an urban context. Speech pathologists, and those interested in speaker characterization for other reasons, such as a focus on forensic phonetics, have addressed the third objective. Speech technology has successfully developed automatic systems for speech production, speaker recognition, and language identification (Laver 1994c); but the methods used mostly exploit automated machine learning about hidden statistical patterns in the acoustic waveforms of speech, which doesn't involve explicit "description" in the same sense. No language investigated so far has been comprehensively

and explicitly described against all four objectives (though general phonetic theory could in principle be applied to each of these tasks).

Within the first objective, linguistic phonetic accounts also often limit themselves to specifying only the phonetic basis for distinguishing the patterns that contrastively identify one phonological unit from another, for example the consonant or vowel phonemes that discriminate minimally different words in English such as *call* and *tall*, or *seal* and *sill*. The contextual patterns associated with the incidence of contrastive linguistic units in different *structures* and in different *environments* are less often described in detail, rich in phonetic regularity though they are. These aspects of sound-patterning often ignored by linguistic phonetic accounts include the wide range of allophonic realizations of phonemes in different syllable structures and in different contextual environments within syllables.

The limiting of linguistic phonetic accounts of languages to a description chiefly of distinctive phonological contrasts is no doubt because it is seen as a means to a different end. An account of phonological contrasts is all that is normally felt by linguists to be needed for further discussion of linguistic behavior at higher levels than phonology, in morphology, lexis, syntax, and semantics. From the phonetician's perspective, however, once these contrastive patterns have been identified, it is in the phonetic detail of the contextual allophonic interaction of linguistic units that some of the most interesting and challenging phenomena in speech production and perception are to be found.

The presentation of a model of linguistic phonetics in this chapter will give priority to describing the phonetic basis for differentiating the contrastive and contextual patterns in speech which signal the identities of linguistic units in the different languages of the world, but will touch in passing on the other objectives as well. The phonetic symbols used in transcription, enclosed in square brackets [ ], will be those of the International Phonetic Alphabet (1993) of the International Phonetic Association (IPA), set out in what is usually called the IPA Chart. The chart is attached as an appendix to this chapter, for consultation about transcriptional symbols and their classificatory phonetic identification.

4 The Coverage of a Linguistic Phonetic Theory

When the full range of the vocal sound-making capabilities of the human species is considered, it becomes apparent that only a restricted subset of the range is used as the basis for contrastive and contextual patterns in spoken language. To offer a few examples, no language makes distinctive use of the percussive noise of the teeth colliding as the jaw is snapped shut. Nor is the noise of air being squeezed between the cheek wall and the outer surface of the teeth and gums used in language by normal speakers (though it is sometimes used as a substitute for the voice by speakers who have had their larynx removed by surgery). The ability to simulate a snoring sound is not used contrastively, nor is a falsetto voice used deliberately to contrast one sound against another, in any known language.

There is a further degree of constraint. Not only is the range of sounds that is used in language limited to a relatively small subset of those physiologically possible, but within that subset there is a core of frequently used sounds that turn up repeatedly in widely different language families, within a broader range of less frequent sounds. As part of that core, most languages use [t], [n] and [s] as consonants, as in the pronunciations of English *tea*, *knee*, and *sea*. Relatively few, on the other hand, use consonants such as the initial sounds [f] in English *fin*, [θ] in *thin* or [ð] in *then*. A very large number use the vowels [i], [a], and [u], as in English *peel*, *pal*, and *pool*. But very few use the vowels [y], [ø] or [œ], as in French *lune* ("moon"), *yeux* ("eyes"), or *peur* ("fear") respectively. Only about one-third of all known languages use diphthongs, such as [aʊ], [eɪ] and [ɔɪ], in the word-final syllables of the English verbs *allow*, *allay*, and *alloy* (Lindau, Norlin, and Svantesson 1990).

There seem to be five interactive principles that may explain this human tendency to use a somewhat restricted number of sound types for purposes of linguistic communication (Lindblom 1983, 1986; Ohala 1989; Stevens 1972). These are:

1 perceptual stability;
2 adequate perceptual contrast;
3 ease and economy of articulatory performance;
4 ecological robustness;
5 ease of modifiability to the needs of the communicative situation.

Perceptual stability is achieved by languages tending to use sounds for which small articulatory adjustments make little auditory difference. Maintaining adequate perceptual contrast entails avoiding sound differences close to the limits of human discrimination. Ease and economy of articulation are the outcome of choosing sound types which do not unduly tax the capabilities of the speech production system. Ecological robustness reflects the ability of sounds to resist the perceptual masking effects of other sounds likely to be heard in the environment (especially speech from other speakers). Finally, given that the relative speed, loudness, and articulatory precision of the speech of a given speaker change frequently in response to variations in the social and physical circumstances of the conversation, it is helpful if parameters of speech control are used which can be appropriately modified without damaging intelligibility.

Different languages, and a given language at different times, reach differing solutions to the trading relationships between these five principles. That these solutions are not always optimal is one potential basis for the sound patterns of languages changing over time.

5 The Shape of a General Phonetic Theory

The obedience of spoken language to the five principles described above has an impact on the desirable shape of a general phonetic theory. A well-designed general phonetic theory is one whose posited features and organizational units cover the maximum range of data with the simplest descriptive constructs. If spoken languages in general tend most frequently to favor a core of speech sounds which are perceptually stable, adequately contrastive, relatively easy to articulate, ecologically robust, and intelligible in variable circumstances, then the basic constructs to be set up in general phonetic theory should be the ones whose nature and relationships give the simplest and most economical account of such sounds. The theory is then completed by adding a minimum set of more elaborate constructs, to cover the less frequent and usually more complex sounds.

6 Organic and Phonetic Aspects of Speech

Within the model of general phonetic theory to be offered here, it will be convenient first to distinguish *organic* versus *phonetic* factors in speech. *Organic factors* are those which are "to do with anatomical structure or morphology, and with the constraints which that structure imposes on the potential for physiological action" (Mackenzie Beck 1997: 256). *Phonetic factors* are those which arise from any learnable aspect of use of the vocal apparatus, with its acoustic and perceptual correlates (Laver 1994a: 28). The interplay between organic and phonetic factors in speech is one of the major sources of acoustic variation between different speakers. The recovery of relatively invariant properties in speech data from different speakers, to aid the decoding of linguistic messages (Perkell and Klatt 1986), can only be achieved by resolving the relative contributions of organic and phonetic factors.

Many theoretical and practical consequences arise from the fact that any two speakers of normal anatomy must be treated as capable of producing phonetically identical utterances, despite the often very substantial organic differences between them. The fact that the vocal organs of different speakers can be of very different sizes means that speech from two individuals can be acoustically very different, in absolute physical terms. Comparability of pronunciation therefore arises from considering not the absolute values of acoustic parameters, but their values relative to the individual speaker's own acoustic potential. So the intonational value of the pitch of a large adult male speaker's voice can be compared to that of a small female child by considering in each case whether the pitch should be counted as high (or mid, or low), in relation to the speaker's own pitch range (Ladd 1996). In absolute terms, the voice pitch ranges of these two speakers would be very unlikely to show any physical overlap at all. In relative terms, however, they can be brought into comparability, and when heard as the same in these terms they can be regarded as phonetically equivalent.

The same situation applies to comparisons of the phonetic quality of different speech sounds. Vowel sounds, for example, are acoustically characterized by patterns of resonant frequencies of the vocal tract (Ladefoged 1993). The absolute values of the resonant frequencies depend on the overall length and shape of the tract. These frequencies change as the organs of the vocal tract manipulate it into different configurations, within organic limits set by individual anatomy. The configurations of two vocal tracts can be thought to be phonetically equivalent when the ratios of the lowest resonant frequency to higher resonant frequencies in each of the two cases are closely similar. In absolute terms, given that the resonant frequency ranges for two such organically different speakers as the large man and the small girl would once again show virtually no overlap, it would not be feasible to say that these two speakers were producing comparable sounds. In relative terms, however, they can both be perceived as producing the same vowel [u:] in their pronunciations of the English word "boot" [bu:t], for instance, when the resonant frequencies of each of them show appropriately similar ratios.

Phonetic equivalence is one endpoint of a more general scale of *phonetic similarity*, which is a metric for comparing the phonetic characteristics of any two sounds. The concept of phonetic similarity is hence a necessary basis for the whole of general phonetic theory. In addition, the view that organically different speakers can produce and perceive phonetically equivalent sounds has profound implications for describing normal use by native speakers. Equally profound are the implications for understanding the articulatory and perceptual processes of spoken language acquisition by infants, foreign-language learning by non-native speakers, and pathological use in speech disorders.

The dimension of phonetic similarity is relevant, finally, not only to comparing speech sounds from all different speakers of normal anatomy, but also to two further situations. The first is as the basis, within a single speaker, for grouping phonetically similar allophonic variants into a single phoneme, as a family of phonetically related sounds fulfilling the same contrastive phonological role. The second applies to decisions about the range of phonetic segment types that can be represented by a given character in alphabetically based writing systems for whole language communities. The decision, for instance, about what speech sounds in different languages are eligible to be written with the letter "r" depends in part on the comparability of the phonetic and perceptual qualities of the candidate sounds concerned.

7 Articulatory, Acoustic, and Perceptual Levels of Description of Speech

Emerging from the discussion in the section above is a second general distinction, between three different aspects of the phonetic description of speech. These are related to the three links in the chain of speech, from the speaker's generation of an utterance, to its transmission through the

air, to its reception by the listener. The first is the *articulatory* level of description, which accounts for the changing configurations and other actions of the speaker's vocal apparatus. The second is the *acoustic* level, which consists of statements about the physical consequences of articulatory actions in terms of vibratory patterns of air molecules within the vocal apparatus and in the air between the speaker and the listener. Finally, the third level of description concerns the *perceptual* impressions of the listener receiving the acoustic information.

The rest of this chapter will focus on phonetic aspects of speech, and will be concerned chiefly but not only with the articulatory level of description.

8 Linear and Nonlinear Units of Speech Organization

The phonetic events that make up the time-course of speech tend to be continuous, with only relatively few steady states or sharply defined breaks that could serve as the boundaries of natural, serial units of speech organization. Obvious natural breaks do occur, however, in two circumstances in the linear production of speech by a single speaker. One is at the beginning and end of a *speaking turn* by one participant in a conversation. The other is at the beginning and end of an individual *utterance*, bounded by silence, within the individual speaking turn. Exhaustively dividing the rest of the stream of speech into a sequence of units smaller than the utterance involves appealing to a number of convenient assumptions. A key traditional assumption is that the continuum of speech can be appropriately handled, for analytic purposes, as if descriptive categories were discrete, not continuous. On this basis, it becomes reasonable to set up smaller scale phonetic constructs such as the feature and the segment.

8.1 The relationship between phonetic segments and phonetic features as units of speech production

Phonetic features are collectively the ingredients of *phonetic segments*. In the minimum case, two segments may differ from each other by the presence or absence of just one phonetic feature. A feature exploited in every human language in this way is the phonetic feature of "voicing." Voicing is caused by vibration of the vocal folds in the larynx. Whether the vocal folds vibrate or not will be determined by the interaction of airflow from the lungs and the tension-states of relevant laryngeal muscles. The word-initial consonant sounds in the two English words *zeal* /zil/ ⇒ [zi:l] and *seal* /sil/ ⇒ [si:l] differ in their voicing state, in that the vocal folds are being made to vibrate in the first case (making [z] a "voiced" segment) and not in the second (making [s] a "voiceless" segment). The transcriptional conventions in the example above are that slant brackets // show the phonemic status of the symbols; "⇒" means "is phonetically pronounced as"; square brackets [ ] show the phonetic status of the pronunciation of the words concerned; and [:] after a segment means that the sound is produced "with relatively longer duration."

While segments can be thought of as linear units following one another sequentially in the chain of speech without interval, features are nonlinear. They can overlap each other in time, and have startpoints and endpoints which do not necessarily align with those of the chain of segments. Phonetic segments, representing phonological vowels or consonants, are temporally anchored in the chain of speech by the co-occurrence and mutual timing of their constituent features.

8.2 Phonetic and phonological features

The constructs of a general phonetic theory should include a supposed *universal set of phonetic features*, whose comprehensive coverage of spoken language remains provisionally true until shown by further research to be inadequate. The general phonetic theory summarized here is

based on these principles (Laver 1994a). It tries to include a set of phonetic features capable of describing the phonetic basis of all phonological contrasts, and of all the contextual patterns of their interaction, so far discovered in the spoken languages of the world. The set of phonetic features proposed in a general phonetic theory would nevertheless, in the ideal, always be larger than the set proposed to cover the languages of the world. This is because of the need to extend its coverage to the differentiation of sociolinguistic communities and the characterization of individual speakers.

It is important to appreciate the difference of technical status between descriptive phonetic features and distinctive phonological features. An example was quoted earlier of the phonetic feature of voicing providing the phonetic basis for a minimal contrast between two consonantal sounds in English, /z/ and /s/. Opportunities for conceptual confusion are rife at this point, in that the presence or absence of "voicing" can be seen in two quite different perspectives. Phonetically, the difference between [z] and [s] as physical speech sounds is described in terms of the presence or absence of vibration of the vocal folds, as mentioned briefly above and described in more detail in Section 10.2 below. To expand on the phonetic example mentioned briefly earlier, /z/ and /s/ as consonants in English are phonologically differentiated by the distinctive presence or absence of a single *distinctive feature*, often represented as +VOICE versus –VOICE. (Capitalization of the name of the feature, with "+" and "–" indicating presence versus absence, is a useful way of distinguishing the status of phonological features from that of phonetic features, which often – potentially confusingly, as in this case – have the same or similar names.)

Viewed as a phonetic feature, "voicing" is part of the descriptive, objective vocabulary of phonetics. Viewed as a distinctive feature, VOICE is part of the formal vocabulary of phonology. The purpose of phonetic features is to describe the articulatory, acoustic, or auditory characteristics of speech sounds as events in the real, physical world, independently of the language concerned. The purpose of distinctive features is to focus on the role of the features as part of a conventional, semiotic code for identifying phonological units particular to a given language. The term "distinctive feature" is thus reserved for use as a contrastive phonological concept.

Part of a phonological interest in distinctive features is the exploration of the degree to which different phonological features fall into putatively *natural classes*, where the members of the class share some phonetic and/or distributional property that distinguishes that class from other classes. This often entails grouping classes into more abstract, superordinate classes, such as the phonological class of "sonorant." This superordinate class is normally taken to include the subordinate classes of English vowels, liquids (such as /r, l/), glides (/j, w/) and nasal stops (/m, n, ŋ/). For further discussion of distinctive features and natural classes, see Chapter 10.

8.3 The phonological syllable

The *syllable* is not identified here as a unit of phonetic description. Many phoneticians have tried to develop a robust definition of the properties of a phonetic syllable, but no objective correlate that would link phonetic performance on a one-to-one basis to the phonological syllable has yet emerged (Laver 1994a: 113–15).

The term "syllable" is perhaps best reserved for use at the phonological level, where it is useful (though not itself unproblematic) for two purposes: for the location of word-identifying patterns of stress; and as an organizing concept for the mutual distribution of vowels and consonants. This organization is reflected in the traditional phonological view that *vowels* are nuclear in the syllable, with all syllables containing one and only one vowel.

Consonants are marginal in syllables, being either syllable-initial or syllable-final. Using "C" to mean "a consonant," and "V" to mean "a vowel," the structure of an English monosyllabic word like "strikes" /straɪks/ would be formulaically represented as CCCVCC. Languages differ in the syllable structures they allow. English allows both *open* and *closed* syllables (that is, syllables without and with, respectively, one or more final consonants), as in /aɪ/ "I" V, /saɪ/ "sigh" CV,

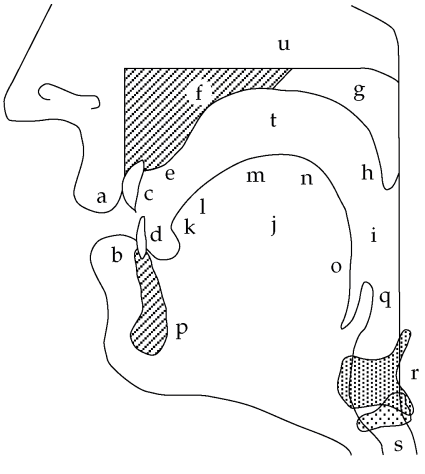
/saɪd/ "sighed" CVC, and /saɪzd/ "sized" CVCC. Hawaiian allows only open syllables, as in the disyllabic word /ola/ "life" V + CV.

9 The Componential Organization of Speech Production

The success of phonetics in developing an objective, replicable, internationally standard method of describing all speech sounds in all spoken languages lies in part in a componential approach to phonetic description. Each discriminable sound is regarded as the composite product of the action of a number of subprocesses of the speech production system. These are described in more detail in Section 10 below. A schematic view of the *vocal organs* which make up the subprocesses, including the *lungs*, the *larynx*, the organs of the *mouth* and the *pharynx* in the vocal tract, and the soft palate (technically called the *velum*), is shown in Figure 9.1.

This componential analysis underlies the conventions of phonetic transcription of the International Phonetic Association. As an illustration of this approach, and to inform the explanation offered below of descriptive phonetic categories, a typical (abbreviated) label for the sound represented in the phonetic transcription of the IPA's International Phonetic Alphabet (1993) as [b] would be "a voiced, labial, oral stop." The four elements of this label constitute individual phonetic features and identify independently controllable components of the production of the sound:

- "*voiced*": the vocal folds in the larynx are vibrating (superimposing aerodynamic pulses on the moving column of air flowing out of the lungs);
- "*labial*": the lips are involved as articulators;



a. Upper lip
b. Lower lip
c. Upper teeth
d. Lower teeth
e. Alveolar ridge
f. Hard palate
g. Velum (soft palate)
h. Uvula
i. Pharynx
j. Body of the tongue
k. Tip of the tongue
l. Blade of the tongue
m. Front of the tongue
n. Back of the tongue
o. Root of the tongue
p. Lower jaw
q. Epiglottis
r. Cartilages of the larynx
s. Trachea (windpipe)
t. Oral cavity
u. Nasal cavity

Figure 9.1 Schematic diagram of a cross-section of the vocal organs. *Source*: After Laver 1994a: 120.

- "*oral*": the velum is in a raised position, sealing off the exit at the back of the mouth to the nasal cavity, causing any airflow to pass through the mouth alone;
- "*stop*": the closure of the lips momentarily seals off the escape of the air in the mouth and pharynx to the outside atmosphere, causing a short-term rise in air pressure in the vocal tract. As the lips open again, the compressed air is then released through them with a small, audible explosion.

An assumption in such abbreviated labeling is that the flow of air is generated by the action of the lungs, with the flow being out of the body. (In a fuller label, this would add an explicit element "with pulmonic egressive airflow" – see Section 10.1.) By identifying the activities of different subprocesses in this way, and with an underlying understanding of the activities thus represented, a componential labeling system in effect offers a set of instructions to informed readers about what to do phonetically with their own vocal apparatus to generate a phonetically equivalent or near-equivalent sound.

10 Speech Production Processes

There are only two basic ways in which a speaker can perceptibly differentiate one segmental speech sound from another – by changing the *phonetic quality* of the sound, or its *duration*. Variation of pitch and loudness play their part at a suprasegmental level when speech is continuous, and may result in differences of meaning, but matters of the prosodic and metrical control of speech production will not be addressed in this chapter. For interested readers, these topics are discussed in Laver (1994a: 450–546), together with issues to do with continuity and rate of speech.

The control of phonetic quality and duration depends on the interaction of five major subprocesses in the production of speech:

1 initiation and direction of airflow;
2 phonation type;
3 articulation;
4 intersegmental co-ordination;
5 temporal organization.

The remainder of this chapter is devoted to an explanation of the way that the activities of these different subprocesses in the control of speech can generate different sounds.

The traditional phonetic approach to the *segmental classification* of speech sounds is said to be a classification by "place and manner of articulation." "Place" will be seen to be straightforward, but "manner" will turn out to be a complex of a range of different types of activity. Segmental classification by place and manner draws on all the factors in the list above, with their interaction producing segments of different phonetic quality and duration.

Description in the sections below will concentrate on the typology of phonetic features, rather than on exemplifying every cell of the resultant matrix of categories. For a comprehensive account of both segmental and suprasegmental categories of speech sounds, the reader is referred to Laver (1994a: 95–546). For the interpretation of specific phonetic symbols, the IPA Chart in the appendix to this chapter should be consulted.

10.1 Initiation and direction of airflow

There are three categories of *initiation of airflow* used for speech, and two of *direction of airflow*. The means of setting a column of air moving can be classified in terms of the initiating mechanism used. By far the most frequent initiator of airflow in speech is the pulmonic mechanism,

setting lung air flowing in an egressive direction to the external atmosphere. This *pulmonic egressive airflow* is then modified in turn by the actions of the larynx, the vocal tract and the velum. Speech made on a *pulmonic ingressive* mechanism, on an inflowing breath, seems to be used only paralinguistically, for example in Scandinavian cultures to express sympathy or commiseration.

The second initiator of airflow used in speech is the larynx, in the *glottalic airstream mechanism*. As the name suggests, the *glottis* (the space between the vocal folds) is involved. With the glottis acting as a valve, and closing off the flow of air from the lungs, the larynx can be abruptly raised or lowered in the throat by muscular action, like a piston in a cylinder. The effect is to compress or rarefy the volume of air in the vocal tract, causing a sharp explosion (on release of a compressed *glottalic egressive* airstream) or an abrupt implosion (on the release of a rarefied *glottalic ingressive* airstream). Sounds made on a glottalic egressive airstream are called *ejectives*, and those on a glottalic ingressive mechanism *implosives*. In Zulu, the word [kʼaːkʼa] "surround" involves two ejectives, symbolized by the apostrophe [ʼ] after the stop symbols.

If voicing is added to an implosive, by pulmonic egressive airflow making the vocal folds vibrate as the larynx descends during a glottalic ingressive initiation, a *voiced implosive* segment is the result. The contrastive difference between the two Hausa words [ɓaɓe] "estrangement" and [babe] "grasshopper" relies on the two stop segments in the first word being voiced labial implosive stops (hence involving two airstream mechanisms, glottalic ingressive and pulmonic egressive), and in the second on the two stop segments being voiced labial pulmonic egressive stops (involving only one airstream).

The third initiator of airflow used in speech is the tongue, in the *velaric airstream mechanism*. Because the tongue is involved, it is therefore also sometimes called the "lingual" mechanism. Velaric sounds are made by the body of the tongue trapping a volume of air between two closures in the mouth, one at the velum, and one further forward. The tongue then retracts the velar closure by sliding backward along the soft palate while maintaining the closed stricture, thus rarefying the air pressure enclosed in the expanded, sealed cavity. When the front closure is then released, the air implodes into the relative vacuum. Sounds made on this *velaric ingressive* airstream are called *clicks*. Since the velaric mechanism is confined to actions within the mouth, the rest of the vocal apparatus is free to add voicing and / or nasality to click sounds.

The languages that use click sounds contrastively are confined to southern and eastern Africa. Ladefoged and Maddieson (1996: 246–80) offer a comprehensive account of these sounds, with many examples from languages such as Nama, Zulu, and Xhosa. In English, clicks are used only paralinguistically, to indicate annoyance (usually written "tsk, tsk" in the English writing system), or to encourage horses to accelerate, or onomatopoeically to simulate the clopping sound of their hooves.

10.2 Phonation type

The biological function of the larynx is chiefly to act as a protective and regulative valve for the airway to and from the lungs. The valving mechanism that has evolved is a delicate and complex muscular structure within a supporting framework of cartilages (Dickson and Maue-Dickson 1982; Laver 1980). The so-called *vocal folds* are two shelves of muscular tissue which run horizontally from front to back of the larynx, capable of separation at the back to leave a flat, triangular space with its apex at the front. This space was identified earlier as the glottis, and there are six modes of *phonation* used in spoken language to distinguish different segments, involving different adjustments of the glottis.

When pulmonic egressive air flows upward from the lungs, a *voiceless* sound is produced if the triangular space of the glottis is left wide open, as if for breathing out. Examples of voiceless consonant sounds widely used in languages are the word-initial sounds in English *see* [siː], *tea* [tiː], and *she* [ʃiː]. If the vocal folds are brought close enough together to make the continuous

airflow through them turbulent, either through a gap left at the back or through a narrowed glottis, the result is called *whisper*.

Voicelessness can be heard in the pronunciation of some vowels in a number of languages. As an allophonic process before pauses (described below as an outcome of the coordinatory process called "devoicing"), French vowels often lose their voicing. An example would be [wi̥:] *oui* ("yes") at the end of an utterance, where [̥] below the symbol indicates voicelessness. Alternatively, this devoicing is often substituted by whisper, rather than strict voicelessness. English also exploits allophonic voicelessness, in optional pronunciations of unstressed vowels between two voiceless consonants, as in the first syllable of *potato*/pəteitəʊ/ ⇒ [pə̥teitəʊ] in Received Pronunciation of British English. Further examples of voicelessness or whisper on vowels in Amerindian, Sudanic, Sino-Tibetan, and Australian languages are given in Ladefoged and Maddieson (1996: 315) and Laver (1994a: 295–7).

In the third type of phonation, vibration of the vocal folds is the basis for *voiced* sounds, as mentioned briefly in Sections 8.1 and 9 above. Examples of voiced sounds widely used in languages are the word-medial consonant sounds [z, d, g, m] in English *easy* [i:zɪ], *aiding* [eɪdɪŋ], *again* [əgeɪn], and *seeming* [si:miŋ], as well as the vowel sounds in these words [i:, ɪ, eɪ, ə]. In voiced sounds, the vocal folds are brought lightly together by muscular action, blocking off the outflow of pulmonic air, and air pressure below the closed folds building until it is sufficient to blow the folds apart against the muscular tension holding them closed. Once airflow is re-established through the glottis, an aerodynamic effect is produced within the glottis, with the egressive pulmonic flow creating very local suction as it passes at high speed through the relatively small gap between the vocal folds. This local force sucks the vocal folds toward each other, and combines with the muscular tension to restore the closed position of the vocal folds. The abrupt restoration of closure sends a small shockwave traveling on the outflowing breath through the vocal tract, and acoustically excites it into resonance.

The cycle from closure of the vocal folds to separation and renewed closure typically happens very fast (in a range from 60 to 240 times per second in adult male voices in normal conversational English). The frequency of the vibration corresponds to the auditory *pitch* of the voice. The contour of pitch in the successive, intermittent voiced sounds of a whole utterance is in effect heard as a melody, and functions as the *intonation* of the utterance.

The fourth type of phonation used in spoken language is *creak* or *creaky voice* ("creak" is also sometimes called *vocal fry* or *glottal fry* in American publications). In this mode of phonation, the front part of the glottis vibrates, at a considerably lower frequency than in normal voicing, while the back part is pressed more tightly together. Pairs of Danish words can be distinguished by the presence of syllable-final creak (sometimes also called *laryngealization*) versus its absence, for instance in [dʊ̰] "tablecloth" versus [du] "you" (Laver 1994a: 330–3).

The fifth type of phonation is *whispery voice* (also sometimes called *breathy voice* or *murmur*). As in whisper, the vocal folds do not completely seal off the transglottal escape of the pulmonic airflow while vibrating, but leave a gap – either at the back of the glottal triangle, or along the length of the approximated but vibrating vocal folds. The result of the continuous leakage of air is to superimpose audible whisperiness on the pulsed voicing throughout the phonation. Whispery voice in English is used phonetically in English as an optional allophonic feature to replace the normal voicing of [h] when that consonant occurs in intervocalic position in some accents of British English, as a whispery voiced resonant [ɦ̤]. Examples are *ahead* /əhɛd/ ⇒ [əɦ̤ɛd] and *perhaps* /pəhaps/ ⇒ [pəɦ̤aps], with phonetic voicing running right through these words, becoming momentarily whispery during the "h". Whispery voice of this sort is also used in English paralinguistically throughout an utterance to signal secrecy or confidentiality.

In a range of other languages, whispery voice is used contrastively to distinguish one consonant phoneme from another. An example is [bəla] "a snake" versus [bɦ̤əla] "good" in Sindhi, using [ɦ̤] in association with the [b] symbol to indicate a whispery-voiced beginning to the syllable in the second word. Section 10.4 below classifies this as an intersegmental co-ordinatory instance of "voiced aspiration" (Laver 1994a: 354).

Finally, closure of the vocal folds may itself constitute the medial phase of a stop segment, in which case it is called a *glottal stop* [ʔ]. Glottal stops are used only allophonically in English, for example as a phonetic realization of the final /t/ consonants in London Cockney *eat that pizza* /iːt ðat piːtsə/ ⇒ [əiʔ ðaʔ pʰəiʔsʌ].

10.3 Articulation

A key part of appreciating how descriptive phonetic classification works is understanding the relationship between segments and features. This section on articulation begins with a clarification of this relationship, and then discusses principles of classification by place of articulation, degree of stricture, multiple degrees of stricture, and aspect of articulation. The technical vocabulary introduced in this section is then used in the discussion of intersegmental co-ordination.

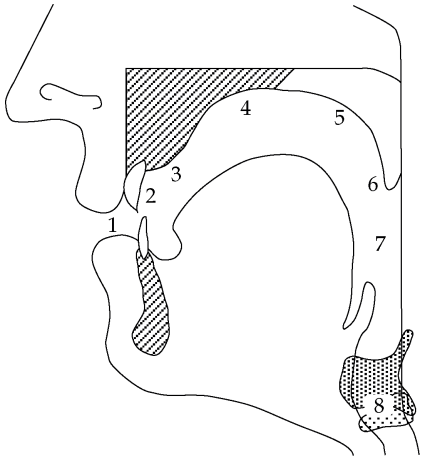
10.3.1 Featural phases of the structure of segments The complex relationship between segments and features can be clarified by appeal to the concept of three internal *phases* of a segment – the onset phase, the medial phase, and the offset phase (Laver 1994a: 112–13). The configuration of the vocal tract during speech changes dynamically from moment to moment between variably greater and lesser degrees of local constriction of the airflow. These constrictions are created by a mobile, *active articulator* (such as the tongue, or lower lip) moving toward a fixed or less mobile, *passive articulator* (such as the hard palate, soft palate, or upper lip). The time occupied in maintaining the maximum degree of articulatory constriction (or degree of *stricture*) reached by the vocal tract during the production of an individual segment delimits the *medial phase* of the segment's performance.

During the *onset phase* of a segment the active articulator is approaching the maximum stricture, and in the *offset phase* is moving away from this toward the configuration for the medial phase of the next segment. One segment's offset phase overlaps with the onset phase of the next segment, in an *overlapping* phase.

The concept of the phasal structure of segments is important for two reasons to do with the temporal distribution of phonetic features. The first is that a given feature may start or finish within a particular segmental phase. For example, in English syllables anticipatory nasality begins relatively early within the medial phase of a vowel segment before a nasal consonant segment, as in *calm* /kam/ ⇒ [kʰãːm]. (In the IPA transcription here, superscript [ʰ] means "aspiration," or "voice onset delay," [˜] means "is nasalized," with the soft palate open to allow airflow into the nasal cavity.)

A given feature may alternatively be coterminous with the medial phase of the segment, as in the case of audible friction in [θ] in English *thin* /θɪn/ ⇒ [θĩn]. Or the feature may run through the medial phases of two or more adjacent segments, as in the English word *soon*, /sun/ ⇒ [sʷũːn], where lip-rounding runs through the first two segments, relaxing to a neutral position toward the end of the word. (The vowel segment [u] is inherently lip-rounded, and in the case of consonant segments lip-rounding is phonetically symbolized by the attachment of the diacritic [ʷ] – see also Section 10.3.4 below on multiple degrees of stricture.) A feature running through adjacent segments can be called a *setting* (Laver 1994a: 115, 391–427), and an analysis of features into settings is useful not only for linguistic phonetics, but also for *paralinguistic* analysis of affective or emotional communication through tone of voice, and *extralinguistic* analysis of speaker characteristics (Laver 1980; Nolan 1983; Pittam 1994).

10.3.2 Place of articulation Classification by *place of articulation* identifies the location of the articulatory zone in which the active articulator is closest to the passive articulator during the medial phase of a segment. An enabling concept for approaching this classification is to distinguish between neutral and displaced places of articulation. In its *neutral configuration*, the vocal tract is as nearly as anatomy allows in equal cross-section along its full length from lips to pharynx.



1. Labial
2. Dental
3. Alveolar
4. Palatal
5. Velar
6. Uvular
7. Pharyngeal
8. Glottal

Figure 9.2 Schematic diagram of some of the neutral places of articulation. *Source*: After Laver 1994a: 135.

(If a vowel sound were to be produced in such a configuration, it would have the quality of the "neutral" vowel [ə] in the pronunciation of the first (unstressed) syllable of the English word *canoe* [kənu]; and acoustically the resonant frequencies would be such that the ratio of the higher frequencies were odd multiples of the lowest.)

In the neutral configuration, the potential active articulators (the lower lip and the tip, blade, front, back, and root of the tongue) lie in their natural anatomical position opposite their passive counterparts along the longitudinal axis of the vocal tract. A segment whose place of articulation is neutral is made by an active articulator moving toward its neutral, passive counterpart. The neutral configuration of the vocal tract, and some labels for *neutral places of articulation*, are given in Figure 9.2.

Neutral places of articulation are thus involved when the bottom lip moves up against the top lip to create a *labial* articulatory narrowing or closure; when the tip of the tongue touches the inner surfaces of the central upper incisors to make a *dental* closure; when the blade of the tongue articulates against the alveolar ridge behind the teeth to make an *alveolar* closure; or when the back of the body of the tongue contacts the soft palate to create a *velar* closure.

When the bottom lip is retracted from its neutral place, however, to articulate instead against the central upper incisors, or the tip of the tongue is retracted to articulate against the alveolar ridge behind the teeth, the resulting *labiodental* and *apicoalveolar* strictures are classified as *displaced articulations*. Similarly, if the blade of the tongue is protruded between the lips, and makes a *linguolabial* closure against the upper lip, that too is a displaced articulation. The labels for neutral and displaced articulations are shown together in Figure 9.3.

Examples of neutral articulations in English are all vowels (except those in some accents where the tongue-tip is curled upward in anticipation of a following /r/), and the word-initial consonant sounds in *pea* [pi:] and *bee* [bi:] (both labial); *theme* [θi:m] and *thee* [ði:] (both dental); *teal* (ti:l] and *deal* [di:l] (both alveolar); *cash* [kaʃ] and *gash* [gaʃ] (both velar); and *he* [hi:] (glottal). Instances of displaced articulations in English are the word-initial consonant sounds in *feel* [fi:l] and *veal* [vi:l] (both labiodental).

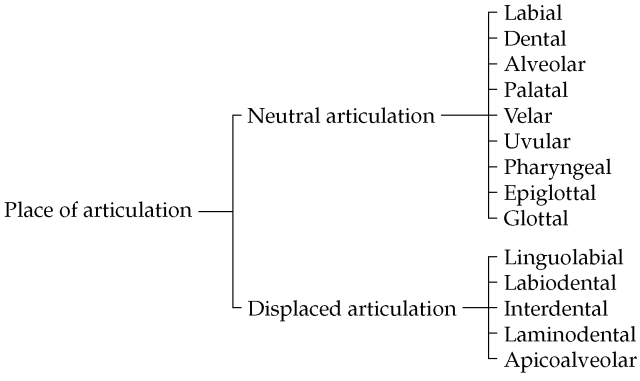


Figure 9.3 Labels for neutral and displaced articulations.

Setting up the classificatory distinction between neutral and displaced articulations amounts to a claim about the relative frequency of incidence of different sounds in the languages of the world. The simpler, less elaborate concept of neutral articulations underpins a broadly sustainable assumption that neutral labial, dental, alveolar, palatal, velar, and glottal sounds are more frequently encountered, for instance, than the displaced linguolabial, labiodental, and apicoalveolar sounds. However, this claim becomes less successful when one considers the relative infrequency in the languages of the world of neutral uvular and pharyngeal sounds, for whose relative rarity more specific reasons would have to be advanced.

10.3.3 Degree of stricture Classification by *degree of stricture* answers the question: "In the medial phase of the segment, to what degree is the gap between the active and passive articulators narrowed?" Languages exploit three types of segments defined by the criterion of degree of stricture – stops, fricatives, and resonants. In the medial phase of *stops*, the degree of stricture is one of *complete articulatory closure*. Examples from English are the word-initial consonant segments [p, b, t, d, k, g] in *post*, *boast*, *toast*, *dosed*, *coast*, and *ghost* respectively.

In *fricatives*, the articulatory stricture in the medial phase is one of *close approximation*, with the airflow made turbulent by passing through a very narrow gap between the active and passive articulators, generating an audible hissing noise ("friction"). Examples of fricatives are the word-initial consonant segments [f, v, θ, ð, s, z] of English *fan*, *van*, *thigh*, *thy*, *sink*, *zinc*, or the word-medial consonant segments [ʃ , ʒ] in English *mesher* and *measure*.

In the medial phase of *resonants* (which can involve sounds representing both consonants and vowels), the stricture is one of *open approximation*. This is a stricture which is sufficiently open to allow the airflow to pass smoothly without turbulence. Open approximation is optimal for allowing the pulses of voiced vibration from the larynx to set the vocal tract into resonance as an acoustic tube.

Examples in an accent of British English of resonants which act as consonants are the word-initial segments representing /j, w, r, l/ in *yield* [jiːld], *wield* [wiːld], *raw* [ɹɔ] and *law* [lɔ]. Examples of resonants acting as vowels from the same accent are the word-final segments in *bee* [biː], *Shah* [ʃaː], *paw* [pɔː] and *two* [tuː]. The IPA chart subclassifies the open-approximation degree of stricture of such resonants in terms of three further articulatory dimensions. The first two are divisions vertically and horizontally of the *vowel-space* in the mouth within which the highest point of the regularly curved tongue is located for the resonant in question. The vertical division is subdivided into *close, close-mid, open-mid*, and *open* resonants. The horizontal division is subdivided into *front, central*, and *back* resonants. The third classificatory dimension for resonants

acting as vowels describes the lip-position of the segment, divided into *rounded* and *unrounded*. The resonant in *bee* [biː] is close, front, unrounded; in *Shah* [ʃaː] is open, back, unrounded; in *paw* [pɔː] is open-mid, back, rounded; and in *two* [tuː] is close, back, rounded.

10.3.4 Aspect of articulation The concept of *aspect of articulation* extends the concepts of "neutral." It is suggested that the majority of stops, fricatives, and resonants in the languages of the world will be performed with the tongue in a regularly curved shape (convex both longitudinally and laterally), with the velum closed, and with a stricture maintained more or less as a steady state throughout the medial phase in a single, neutral place of articulation. This set-up will be treated as a neutral reference against which three non-neutral groups of aspects of articulation can be described. These are the conformational, topographical, and transitional aspects (Laver 1994a: 140–7).

The *conformational aspects* deal with the routing of the airflow channel. There are three distinctions to be drawn. The first is between *oral* airflow versus *nasal airflow*. The second is between *central* versus *lateral airflow*. The third is between *single* versus *multiple strictures*.

As instances of differences between oral and nasal sounds, neutral voiced *oral stops* include [b, d, g], as in English *bib* [bɪb] (*oral labial stops*), *did* [dɪd] (*oral alveolar stops*) and *gig* [gɪg] (*oral velar stops*) respectively. Their non-neutral *nasal stop* counterparts are [m, n, ŋ], as in English *mum* [mʌm] (labial), *none* [nʌn] (alveolar) and *sung* [sʌŋ] (velar). An allophonic difference between an oral and a *nasal fricative* at the same place of articulation is in Igbo "to wedge in" [ɪfa] versus "to shriek" [ĩfã] (Williamson 1969: 87), from Nigeria. Here both are non-neutral in a different respect, in that they share a displaced labiodental place of articulation. A phonemic difference between an oral and a *nasal resonant* can be found in Sioux "sun" [wi] versus "woman (abbreviated form)" [wĩ] (J. Harris, personal communication).

Stop articulations can show complex aspectual patterns of oral and nasal sequences within the medial phase of a stop. The place of articulation of the oral and nasal elements are *homorganic* – the oral stricture is at the same place of articulation. When the nasal element is minor compared with the duration of fully oral closure, and occurs at the beginning of the medial phase, the stop is said to be *prenasal*; when it is final with respect to the oral closure, it is called a *postnasal stop*. The duration of such nasal elements is shorter than in full segmental sequence of nasal + oral stops, as in English *candor*, for instance. When the nasal element dominates the duration of the oral closure in the medial phase, it is said to be a *preoccluded* or *postoccluded nasal stop*, depending on the initial or final location of oral closure. Examples of *complex oral / nasal stops* are found in a range of languages, including some in Africa, India, and South and Central America. An instance of prenasal stops comes from Kalam, a Papuan language of New Guinea, in "down valley" [ᵐbim] and "sinew" [kiⁿdɨl] (Pawley 1966). A fuller discussion of such complex oral / nasal stops is offered in Laver (1994a: 227–35).

For the sake of economy, sounds will from now on be assumed to be oral unless specific mention is made of their nasal status.

In the case of differences between central and lateral sounds, a neutral example would be the *voiceless alveolar central fricative* [s], as in English *sea* [siː]. A non-neutral instance would be a *voiceless alveolar lateral fricative* [ɬ], as in North Welsh "her ship" [iɬɔŋ], which is in phonemic contrast with a *voiced alveolar lateral resonant* "his ship" [ilɔŋ] (Albrow 1966: 2). In both lateral cases, the air flows through a gap at one or both sides of the tongue behind a central contact between the tip or blade of the tongue against the alveolar ridge.

The active articulators of the vocal tract are sufficiently flexible and versatile to be able to create articulatory strictures in two different places simultaneously (i.e., sharing the same medial phase). When two such strictures are of equal degree the conformational aspect of articulation shows (non-neutral) *double articulation*. Two examples from the West African language Yoruba are the words [k͡pe] "to call" and [g͡be] "to carry" (Bamgbose 1969: 164). The two simultaneous closures in these double stops [k͡p] and [g͡b] are made at the labial and velar places of articulation, and they are therefore called (voiceless and voiced) *labial velar stops*.

An example of a double articulation involving a consonant sound in English is the initial segment in *well* [wɛl], in which the lips are in a rounded position, and the back of the tongue is raised to a position close to the soft palate, but in neither case close enough to create local friction. The result is a *labial velar resonant*.

When one stricture is of greater degree than the other during the shared medial phase, the narrower stricture is said to be the *primary articulation*, and the more open stricture is called the *secondary articulation*. The auditory effect of secondary articulations is usually to add a modifying "coloring" to the perceptual quality of the primary articulation. Examples include *labialization*, which adds lip-rounding to a segment; *palatalization*, in which the front of the tongue is raised toward the hard palate; *velarization*, in which the back of the tongue is raised toward the soft palate; *pharyngealization*, in which the root of the tongue is retracted toward the back wall of the pharynx; and *nasalization*, in which the soft palate is lowered, allowing air to flow through the nasal cavity and add nasal resonance to the oral resonance of the rest of the vocal tract.

The auditory effect of a neutral, single stricture without secondary articulation is sometimes referred to as "plain." The quality associated with palatalization is sometimes said impressionistically to be "clear," and that with velarization and pharyngealization "dark." In most accents of English, there is a structural allophonic difference between the pronunciations of /l/ in syllable-initial position and in syllable final position, in that both show secondary articulations, with the /l/ of *leaf* [lʲi:f], for instance, being a ("clear") *palatalized voiced alveolar lateral resonant* and that of *feel* [fi:ɬ] a ("dark") *velarized voiced alveolar lateral resonant*. Another English example of secondary articulation is the *labialized palatoalveolar fricative* initial in *she* [ʃʷi:], where the primary articulation is the fricative stricture midway between the alveolar and palatal places of articulation (hence "palatoalveolar," symbolized by [ʃ]), and the secondary articulation is one of rounding of the lips, symbolized by the superscript diacritic [ʷ]). The use of secondary articulation is discussed further in the section on intersegmental co-ordination below. Figure 9.4 summarizes the labels for double and secondary articulation.

The *topographical aspects* deal with the shape of the tongue as the active articulator both longitudinally and transversely. Laver (1994a: 141–2) discusses longitudinal processes such as extending or withdrawing the tip of the tongue, and advancing or retracting the root of the tongue. But the most frequently found topographical aspect involving the long axis of the tongue is *retroflexion*, in which the tongue tip is curled up and backward, sometimes to the extent of presenting the underside of the tip to the roof of the mouth. Margany, a language of South Queensland, shows a phonemic contrast between a (neutral) voiced alveolar stop and a *voiced postalveolar retroflex stop*, in "to cry" [badi] and "maybe" [baɖi] respectively (Breen 1981).

The major transverse aspect distinguishes a flat blade of the tongue from one in which the blade is *grooved*. In English, [s] is produced by most speakers with the air flowing through a very narrow channel in the tongue along the surface of the blade just opposite the alveolar ridge. *Flat alveolar fricatives* occur in Icelandic. In this case it is not clear which category should be treated

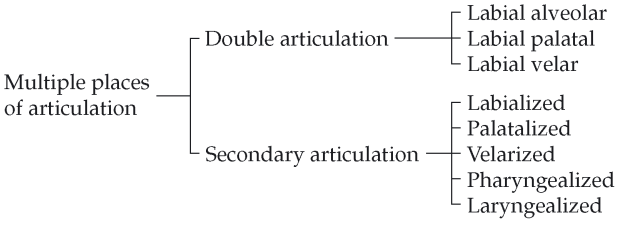


Figure 9.4 Labels for double and secondary articulations.

as neutral, in that grooved alveolar fricatives are far more common than flat alveolar fricatives. It is possible that the higher "pitch" of the fricative noise made through a narrow groove is more audible, and hence ecologically more robust, than the lower pitch of a fricative made through a broader, flat gap. But the articulatory adjustment for creating a central groove is physiologically more complex than for a flat gap, and it may be that the concept of a neutral articulation as the more natural and more widespread sound breaks down at this point.

The *transitional aspects* handle the question of whether the active articulator is static during the medial phase of the articulation, or in dynamic movement. In performing a neutral stop articulation such as the voiced alveolar stop [d], the blade of the tongue rises at moderate pace up toward the alveolar ridge as the passive articulator, makes contact for an appreciable duration, then descends. A (non-neutral) *voiced alveolar tapped stop* is like the neutral version, but moves much faster into contact, makes a very brief closure with the alveolar ridge, and moves away fast. An example is found in many American English accents, as the pronunciation for "t" between two vowels, in a word like *city* [sɪɾɪ].

A tapped stop is sometimes likened to one tap of a trilled stop, another non-neutral example. A trilled stop is one where the active articulator, such as the tip of the tongue, is positioned close to the passive articulator and the airflow through the narrow gap (analogous to the aerodynamic situation in voiced vibration of the vocal folds) brings it repeatedly into full contact. The symbol for a *voiced oral alveolar trilled stop* is [r], and for one made at the uvular location is [ʀ]. A language that contrasts voiced alveolar tapped and trilled stops is Kurdish, as in the pair of words "wound (injury)" [bɾin] versus "cutting" [brin] (A. Ferhardi, personal communication). A contrast between an alveolar tapped stop and a uvular trilled stop is found in European Portuguese, in "dear" [karu] versus "car" [kaʀu] (Parkinson 1988: 138).

When a stop is *flapped*, it strikes the passive articulator in passing. A (non-neutral) *voiced oral alveolar retroflex flapped stop* [ɽ] starts with the tongue-tip curled upward, and then in uncurling the tip strikes the alveolar ridge very briefly, making a sliding contact that is quickly broken. Westermann and Ward (1933: 76) cite the Sudanese language Gbaya as contrasting a trilled stop with a flapped stop, in "beans" [ere] versus "hen" [eɽe].

Transitional aspects of articulation affect resonants as well. A *monophthong* is phonetically a (neutral) resonant segment with a relatively steady-state articulatory position being maintained throughout its medial phase. A *diphthong* is a (non-neutral) resonant which changes its articulatory position from one position of open approximation toward another during the medial phase. A *triphthong* is a (non-neutral) resonant which changes articulatory position during the medial phase from one position of open approximation toward another and then another. English is unusual amongst the languages of the world in that resonants acting as vowels can show all three transitional aspects of articulation. In some accents of British English, the vowel in a syllable may be represented by either a monophthong (as in *awe* [ɔ]), a diphthong (as in *eye* [aɪ]) or a triphthong (as in *ire* [aɪə]).

10.4 Intersegmental coordination

Segmental description in this chapter so far has been limited to events within the boundaries of a single segment. Some of the most phonetically interesting events occur in the overlapping phase between two adjacent segments, where the first segment's offset phase is coterminous with the next segment's onset phase (Laver 1994a: 339–90). When a segment is next to utterance-marginal silence, the onset and offset phases involve transitions from and to the articulatory rest position. Also relevant is the effect of the characteristics of one segment's medial phase spreading, anticipatorily or perseveratively, into part or all of the medial phase of the adjacent segment. Significant phonetic events involving coordination of adjacent segments include the phenomena of devoicing, aspiration, release, affrication, and coarticulation.

10.4.1 Aspiration When a segment that in most contexts is fully voiced throughout its medial phase occurs next to a silent pause, say in utterance-initial position, the transition from silence may have the effect of delaying the beginning of voicing for that segment. In most accents of English, voicing for an utterance-initial voiced stop or fricative will start after beginning of the medial phase. Using [#] to indicate silence, and a subscript [ˬ] to mean "delay in the onset of voicing" the utterance-initial word "zeal" would be transcribed [#ˬ zi:l]. Because [z] in most contexts is normally pronounced with full voicing through its medial phase, this delay is usually referred to as *initial devoicing* (though the phonological orientation of such a practice should be noted). Correspondingly, when such a segment is next to utterance-final silence, there may be an early offset of voicing, and the sound is said to be *finally devoiced*. Such utterance-final devoicing would be transcribed, in the English word "lees," as [li:z ˬ #].

In both the initial and final cases, the devoicing is partial, in that not all of the medial phase is deprived of vibration of the vocal folds. When there is no voicing at all in the medial phase, the question is prompted of what differentiates a *fully devoiced segment* such as [z̥] from its voiceless counterpart [s]. Some phoneticians and phonologists make appeal to issues of differential muscular tension in the vocal apparatus, and set up the categories of *lax* and *tense* to describe hypothesized factors that continue to differentiate such devoiced and voiceless segments. It is probably more satisfactory, at a phonetic level of description, to accept the nondifferentiability of fully devoiced and voiceless segments. Figures 9.5a and 9.5b characterize the timing relationships

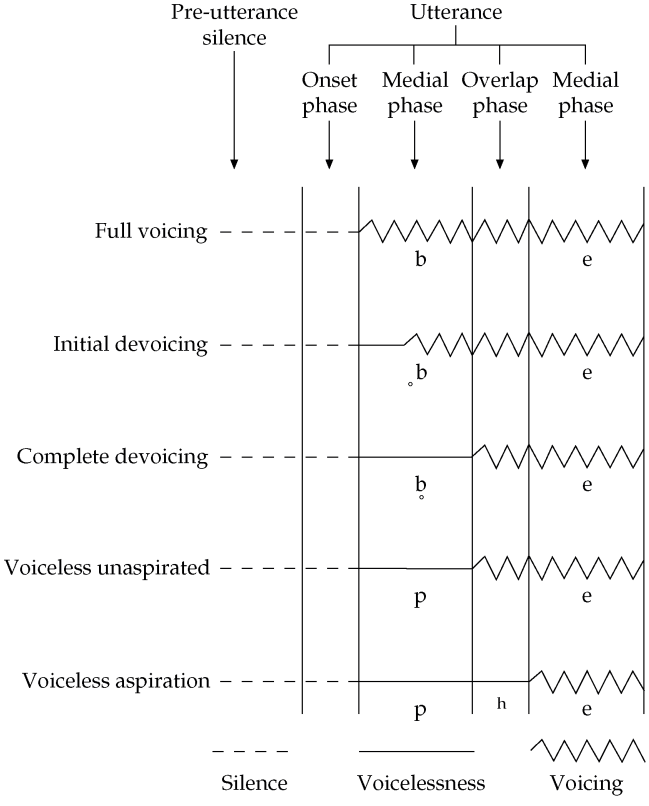


Figure 9.5a iming relationships between laryngeal and supralaryngeal events in initial devoicing and aspiration. *Source*: After Laver 1994a: 340.

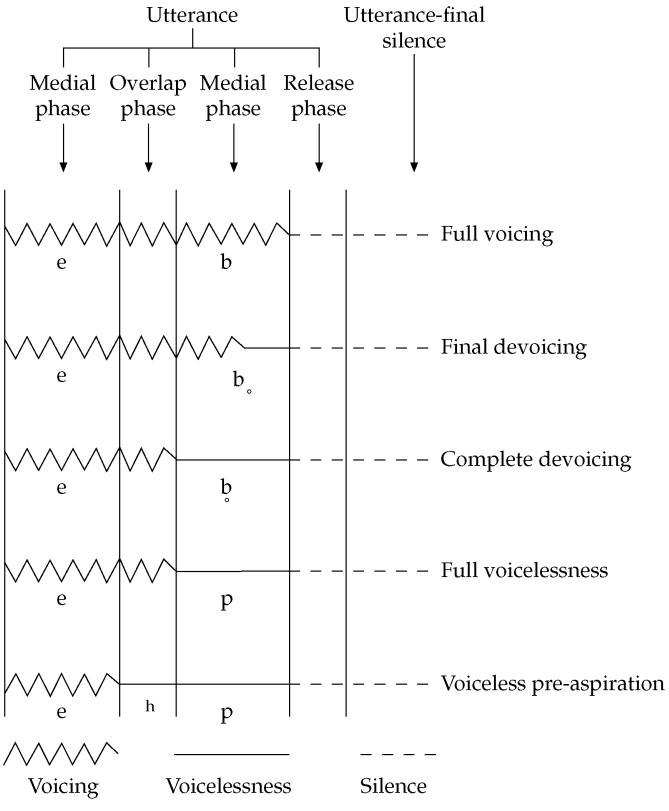


Figure 9.5b Timing relationships between laryngeal and supralaryngeal events in final devoicing and preaspiration. *Source*: After Laver 1994a: 341.

between the laryngeal and supralaryngeal events in the devoicing process, and relate them to the next category of co-ordination to be discussed, aspiration.

When a voiceless segment such as an oral stop is initial before a resonant in a stressed syllable in most accents of English, there is an audible delay in the onset of voicing after the end of the stricture of the medial phase, in the overlap phase between the stop and the resonant. This phenomenon is called *aspiration*. An instance is the English word "peat" [pʰi:t], where the aspiration is transcribed as a small superscript "h." The audible quality of the [ʰ] anticipates that of the oncoming resonant, for which the vocal tract is already assuming the relevant articulatory position. Aspiration is reasonably rare among the languages of the world. French, for example, does not aspirate syllable-initial stops in such circumstances, in words such as "paté" [pate]. Aspiration acts as an allophonic process in English, applying to all voiceless stops /p, t, k/, but is exploited phonemically in a number of languages, including Chengtu Szechuanese, in words such as "to cover" [kai] versus "to irrigate" [kʰai] (Fengtong 1989: 64).

Aspiration is perhaps best defined as "a delay in the onset of normal voicing," since a category of *voiced aspiration* is found in a number of languages of the Indian subcontinent and in central and southern Africa, as a relationship between voiced stops and following resonants. In this case, the phonatory quality of the transition from the stop to the following resonant is one of whispery voice, usually becoming normal (i.e., without audible glottal friction) before the end of

the resonant. Examples of voiced aspiration, involving both oral and nasal stops, are the Sindhi words "to speak ill of others" [gɪla] versus "wet" [gɦɪla], and "in" [mẽ] versus "a buffalo" [mɦẽ] (Nihalani 1975: 91).

Parallel to aspiration as a late onset of voicing in syllable-initial contexts is early offset of voicing in a resonant before a voiceless segment in syllable-final position, which is called *pre-aspiration*. This is a characteristic of many of the circumpolar languages. It can be voiceless, as in an example from Icelandic in "thank" [θahka] (Ewen 1982), or voiced, as in Hebridean Gaelic (of Lewis) in "bag" [pʰɔɦk] (Shuken 1984: 127).

10.4.2 Stop release The offset phase of any oral stop may release the compressed air built up during the medial phase in a variety of ways. Alternatively, the stop may be incomplete, and lack a final release. In this latter situation, the oral closure is sometimes reinforced by a simultaneous glottal stop, as a double articulation. Both modes are found in English, as optional variants. A syllable-final *unreleased stop* can be transcribed for an accent of British English as *top* [tɒpˀ] (and the glottally reinforced version as [tɒp͡ʔˀ]), with the released version as [tɒpʰ].

When the offset phase of a syllable-final stop is released, the channel for the release may be either oral or nasal, central or lateral. In English, the word *button* can be pronounced with *oral release* as [bʌtən], or *nasal release* as [bʌtn̩]. In the latter case, the diacritic below the [n̩] symbol indicates that the nasal segment is "syllabic," acting as the nucleus of the second syllable. The difference between central release and lateral release can be seen in the English word *little*, pronounced with *central release* as [lɪtɔl], or with *lateral release* as [lɪtl̩].

10.4.3 Affrication Affrication is also a characteristic of the offset, release phase of stops. The stop closure is released more slowly than in a nonaffricated stop, so that a brief moment of audible friction is heard as the stricture passes momentarily from complete closure through close approximation. In English, the sounds at the beginning of the words *cheap* [t͡ʃiːp] and *jeep* [d͡ʒiːp] are *affricated stops* (or "affricates"). The special relationship between the stop and the fricative element is symbolized by the use of the linker diacritic. The fricative element of an affricate is by definition homorganic with the stricture of the associated stop element, and affricates can be made at any place of articulation where stops can be formed. The fricative element may also be lateral or central, as in the first and second affricates respectively in the Nahuatl phrase "sit down please!" [ʃimot͡ɬali t͡sino] (Suárez 1983: 32).

10.4.4 Coarticulation and assimilation It is not surprising, given the rate that segments follow each other in the stream of speech, that one segment may influence the articulatory characteristics of segments yet to be spoken, or be influenced by those that precede it. When such an influence crosses a word boundary, it is said to result in assimilation; when it is restricted to word-internal action, it is said to show coarticulation.

An example of *perseverative assimilation* in an accent of British English is the continuing influence of voicelessness across the word boundary in *what's this?* pronounced as [wɒts ðɪs], where the underlying /z/ of *is* loses its voicing under the influence of the preceding [t]. An instance of *anticipatory assimilation* is *that boy* pronounced as [ðap bɒɪ], where the place of articulation of the underlying /t/ is made identical to that of the following [b].

Anticipating the strictural requirements of oncoming segments in the stream of speech often results in secondary articulations. Labialized consonant sounds are found in English before vowels which have rounded lip-positions, in words such as [pʷuːl] *pool*, [mʷɔː] *maw*, and [fʷʊl] *full*. A further example of a secondary articulation that characteristically anticipates future segmental requirements is nasalization. In *anticipatory nasalization*, when a vowel sound precedes a nasal consonant, the soft palate opens during the medial phase of the resonant, anticipating the oncoming requirement for nasality. A language-differentiating facet of this process is that in these circumstances, the soft palate opens later in French than in English, presumably because

of the need to protect the perceptual distinctiveness of French nasal vowel sounds in phonemic opposition to their oral counterparts.

Coarticulatory *anticipations of place of articulation* also occur. In English, the stricture of velar stops such as [k] before a front resonant as in *keep* [kʰi:p] is made further toward the front of the mouth than [k] before a back resonant as in *calm* [kʰa:m].

10.5 Temporal organization of speech

The discussion so far has concentrated on matters to do with the phonetic quality of speech sounds. The remaining variable is *duration*. Segments have certain inherent durational constraints which have physiological or perceptual explanations (Laver 1994a: 431–6). This section will concentrate, however, on the contrastive and contextual control of duration for phonological purposes. The terms "length," "long," and "short" will be reserved for use at a contrastive level, and greater or less "duration" for use at the phonetic level of description.

Phonemic distinctions of length in both vowels and consonants have been observed, with vowel-length distinctions predominating. Vowel-length distinctions abound in accents of English, though usually with associated differences of segment quality. An example of a language using *contrastive vowel length* (with length signaled by the diacritic [:]) is Rarotongan Māori, in word pairs such as "taro bed" [paʔi] and "ship" [paːʔɪ] (Buse 1966: 52).

Phonemic distinctions of consonant length are much rarer, but are found occasionally, as in the Eskimo-Aleut languages of the Canadian Arctic. Inuktitut (Inuit) distinguishes short and long consonants in phrases such as "they arrive together" [tikiqataujutʰ] versus "they arrive frequently" [tikiqatːaqtutʰ] (Esling 1991).

An instance of a language (unusually) contrasting both vowels and consonants is Finnish. An example of such a word pair is "a crease" [rʏpːʏ] versus "a drink" [rʏːpːʏ] (T. Lauttamus, personal communication).

Allophonic adjustments of duration both to structural position and to phonetic environment are very common. In English, the duration of vowel sounds is greatest in open syllables such as *bee*, and less in closed syllables such as *beat*. In syllables of comparable structure, the duration of a vowel sound is greater before a voiced consonant segment such as [d] in *bead*, and less before a voiceless consonant segment such as [t] in *beat*.

11 Conclusion

The detailed resources of general phonetic theory that have only been able to be sketched in here are probably adequate for the task of describing the segmental makeup of almost all languages known today, though of course some problems of detail remain to be resolved. In the terms introduced at the beginning of this chapter, general phonetic theory is basically fit for the linguistic phonetic purpose of "describing the phonetic basis for differentiating all contrastive and contextual patterns in speech which signal the identities of linguistic units in any given language." Not so evident is whether it is yet fit for the three other purposes identified as relevant to a broader interest in linguistic phonetics – the description of phonetic regularities in the speech styles of sociolinguistic communities, of the characteristic phonetic events that distinguish the speech patterns of individual members of those communities, and of the ways in which languages sound different from each other. Such questions raise large issues for the future about the nature and motivation of work in phonetics, and about the desirable and useful limits of resolution of the descriptive apparatus used.

REFERENCES

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Albrow, K. H. (1966). Mutation in spoken north Welsh. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins (eds), *In Memory of J. R. Firth*. London: Longmans Green.

Bamgbose, A. (1969). Yoruba. In E. Dunstan (ed.), *Twelve Nigerian Languages: A Handbook on their Sound Systems for Teachers of English* (pp. 163–72). London: Longmans Green.

Breen, J. G. (1981). Margany and Gunya. In R. M. W. Dixon and B. J. Blake (eds), *Handbook of Australian Languages*, Vol. 2 (pp. 274–393). *Wargamay, the Mpakwithi Dialect of Anguthimri, Watjarri, Margany and Gunya Tasmanian*. Canberra: Australian National University Press.

Buse, J. E. (1966). Number in Rarotongan Maori. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins (eds), *In Memory of J. R. Firth*. London: Longmans Green.

Catford, J. C. (1977). *Fundamental Problems in Phonetics*. Bloomington, IN: Edinburgh University Press and Indiana University Press.

Catford, J. C. (1988). *A Practical Introduction to Phonetics*. Oxford: Clarendon Press.

Catford, J. C. (1994). Articulatory phonetics. In R. E. Asher and J. M. Y. Simpson (eds), *The Encyclopedia of Language and Linguistics*, Vol. 6 (pp. 3058–70). Oxford: Pergamon.

Clark, J. and Yallop, C. (1995). *An Introduction to Phonetics and Phonology*, 2nd edn. Oxford: Blackwell.

Coulmas, F. (1992). *Language and Economy*. Oxford: Blackwell.

Dickson, D. R. and Maue-Dickson, W. (1982). *Anatomical and Physiological Bases of Speech*. Boston, MA: Little, Brown.

Esling, J. (1991). *Phonetic Database (Instruction Manual)*. Pine Brook, NJ: Kay Elemetrics Corporation and University of Victoria, British Columbia: Speech Technology Research.

Ewen, C. J. (1982). The internal structure of complex segments. In H. van der Hulst and N. Smith (eds), *The Structure of Phonological Representations*, Vol. 2 (pp. 26–67). Dordrecht: Foris.

Goldsmith, J. A. (1995). *The Handbook of Phonological Theory*. Oxford: Blackwell.

Fengtong, Z. (1989). The initials of Chengdu speech. *Journal of the International Phonetic Association*, 15, 59–68.

Hardcastle, W. J. and Laver, J. (eds) (1997). *The Handbook of Phonetic Sciences*. Oxford: Blackwell.

Kent, R. D. and Read, C. (1992). *The Acoustic Analysis of Speech*. San Diego, CA: Singular Publishing Group.

Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladefoged, P. (1971). *Preliminaries to Linguistic Phonetics*. Chicago, IL: University of Chicago Press.

Ladefoged, P. (1993). *A Course in Phonetics*, 3rd edn. New York: Harcourt Brace Jovanovich.

Ladefoged, P. (1997). Linguistic phonetic descriptions. In W. J. Hardcastle and J. Laver (eds), *The Handbook of Phonetic Sciences* (pp. 589–618). Oxford: Blackwell.

Ladefoged, P. and Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Laver, J. (1994a). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Laver, J. (1994b). Speech. In R. E. Asher and J. M. Y. Simpson (eds), *Encyclopedia of Language and Linguistics*, Vol. 8 (pp. 4101–9). Oxford: Pergamon.

Laver, J. (1994c). Speech technology: an overview. In R. E. Asher and J. M. Y. Simpson (eds), *Encyclopedia of Language and Linguistics*, Vol. 8 (pp. 4274–89). Oxford: Pergamon.

Lindau, M., Norlin, K., and Svantesson, J. (1990). Cross-linguistic differences in diphthongs, *Journal of the International Phonetic Association*, 20, 10–14.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (ed.), *The Production of Speech* (pp. 217–46). New York: Springer.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala and J. J. Jaeger (eds), *Experimental Phonology* (pp. 13–44). Orlando, FL: Academic Press.

Mackenzie Beck, J. (1997). Organic variation of the vocal apparatus. In W. J. Hardcastle and J. Laver (eds), *The Handbook of Phonetic Sciences* (pp. 256–97). Oxford: Blackwell.

Nihalani, P. (1975). Velopharyngeal opening in the formation of voiced stops in Sindhi. *Journal of the International Phonetic Association*, 32, 89–102.

Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.

Ohala, J. (ed.) (1989). *On the Quantal Nature of Speech*. Theme issue of *Journal of Phonetics*, 17.

Pawley, A. (1966). The structure of Kalam: a grammar of a New Guinea Highlands language. PhD dissertation, University of Auckland.

Perkell, J. S. and Klatt, D. H. (1986). *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Erlbaum.

Pittam, J. (1994). *Voice in Social Interaction: An Interdisciplinary Approach*. Thousand Oaks, CA: Sage.

Shuken, C. (1984). [ʔ], [h] and parametric phonetics. In J. A. Higgs and R. Thelwell (eds), *Topics in Linguistic Phonetics in Honour of E. T. Uldall*. Occasional Papers in Linguistics and Language Learning, no. 9. Coleraine: Dept of Linguistics, New University of Ulster.

Stevens, K. N. (1972). The quantal nature of speech: evidence from articulatory-acoustic data. In E. E. David and P. B. Denes (eds), *Human Communication: A Unified View* (pp. 51–6). New York: McGraw-Hill.

Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Suárez, J. A. (1983). *The Mesoamerican Indian Languages*. Cambridge: Cambridge University Press.

Westermann, D. and Ward, I. C. (1933). *Practical Phonetics for Students of African Languages*. London: Oxford University Press.

Williamson, K. (1969). Igbo. In E. Dunstan (ed.), *Twelve Nigerian Languages: A Handbook on Their Sound Systems for Teachers of English* (pp. 85–96). London: Longmans Green.

APPENDIX

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks		Voiced implosives		Ejectives	
ʘ	Bilabial	ɓ	Bilabial	ʼ	as in:
ǀ	Dental	ɗ	Dental/alveolar	pʼ	Bilabial
ǃ	(Post)alveolar	ʄ	Palatal	tʼ	Dental/alveolar
ǂ	Palatoalveolar	ɠ	Velar	kʼ	Velar
ǁ	Alveolar lateral	ʛ	Uvular	sʼ	Alveolar fricative

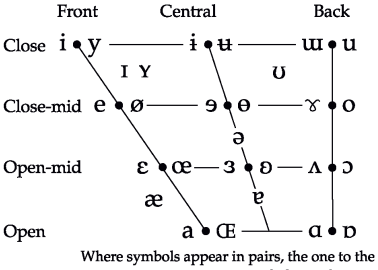
SUPRASEGMENTALS

ˈ	Primary stress	ˈfoʊnəˈtɪʃən	
ˌ	Secondary stress		
ː	Long	eː	
ˑ	Half-long	eˑ	
˘	Extra-short	ĕ	
.	Syllable break	ɹi.ækt	
ǀ	Minor (foot) group		
‖	Major (intonation) group		
‿	Linking (absence of a break)		

TONES & WORD ACCENTS

LEVEL			CONTOUR	
e̋ or	Extra high	ě or	◿	Rising
é	˦ High	ê	◹	Falling
ē	˧ Mid	ě	◿	High rising
è	˨ Low	è	◿	Low rising
ȅ	˩ Extra low	ȇ	◿	Rising-falling etc.
↓	Downstep	↗		Global rise
↑	Upstep	↘		Global fall

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

DIACRITICS

Diacritics may be placed above a symbol with a descender, e.g. ŋ̊

◌̥	Voiceless	n̥ d̥	◌̤	Breathy voiced	b̤ a̤	◌̪	Dental	t̪ d̪
◌̬	Voiced	s̬ t̬	◌̰	Creaky voiced	b̰ a̰	◌̺	Apical	t̺ d̺
◌ʰ	Aspirated	tʰ dʰ	◌̼	Linguolabial	t̼ d̼	◌̻	Laminal	t̻ d̻
◌̹	More rounded	ɔ̹	◌ʷ	Labialized	tʷ dʷ	◌̃	Nasalized	ẽ
◌̜	Less rounded	ɔ̜	◌ʲ	Palatalized	tʲ dʲ	◌ⁿ	Nasal release	dⁿ
◌̟	Advanced	u̟	◌ˠ	Velarized	tˠ dˠ	◌ˡ	Lateral release	dˡ
◌̠	Retracted	i̠	◌ˤ	Pharyngealized	tˤ dˤ	◌̚	No audible release	d̚
◌̈	Centralized	ë	◌̴	Velarized or pharyngealized	ɫ			
◌̽	Mid-centralized	e̽	◌̝	Raised	e̝	(ɹ̝ = voiced alveolar fricative)		
◌̩	Syllabic	n̩	◌̞	Lowered	e̞	(β̞ = voiced bilabial approximant)		
◌̯	Non-syllabic	e̯	◌̘	Advanced Tongue Root	e̘			
◌˞	Rhoticity	ə˞	◌̙	Retracted Tongue Root	e̙			

OTHER SYMBOLS

ʍ	Voiceless labial-velar fricative
w	Voiced labial-velar approximant
ɥ	Voiced labial-palatal approximant
ʜ	Voiceless epiglottal fricative
ʢ	Voiced epiglottal fricative
ʡ	Epiglottal plosive
ɕ ʑ	Alveolo-palatal fricatives
ɺ	Alveolar lateral flap
ɧ	Simultaneous ʃ and x

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.

k͡p t͡s

10 Phonology*

Sound Structure

ABIGAIL C. COHN

1 Introduction

Consider the "words" shown in (1):

(1)		I	II	III	
	a.	*xoda [x] as in German **ch**	poda	poda[z]	(cf. coda, codas)
	b.	*ngatus =[ŋ]	matus	matus[ɪz]	(cf. mattress, mattresses)
		*rudih	hurid	hurid[z]	(cf. herd, hurds)
	c.	*bnick	blick	blick[s]	(cf. block, blocks)

Fluent speakers of English would agree that none of these are actual words of English, yet most speakers would also agree that those in column I are not possible words (we use an * to indicate an impossible or "ungrammatical" form), while those in column II are. In addition, most speakers would agree that the plurals of the would-be words in column II would be pronounced as indicated in column III. How do we know this? Our knowledge of the sound patterns of our native language(s) comes not through memorizing a list of words, but rather by internalizing information about the allowed and disallowed sound patterns of that language. As fluent speakers of English, we know which sounds occur in our language and which don't. For example, in (1a), the [x] sound of German (written **ch** in borrowings from German, as in the German pronunciation of ***Bach***) just doesn't occur in English. In addition, some occurring sounds of English are nevertheless restricted in the position where they occur within the word. As shown in (1b), the sound represented by the spelling sequence **ng** [ŋ] cannot occur at the beginning of a word

* Thanks to Beverley Goodman, Lisa Lavoie, Peggy Renwick, Ayako Tsuchida, and Draga Zec for providing helpful input on earlier drafts of this chapter. Thanks also to Emma Lantz for assisting in the formatting and editing of this version.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

c10.indd 185 12/14/2018 6:33:48 PM

(though it occurs in the middle (*si**ng**er*) or end (*si**ng***)), while **h** cannot occur at the end of a word (but it occurs at the beginning (***h**ot*) or middle (*a**h**ead*)). We also know which sounds can be combined into a sequence. Thus in (1c), **bl** is an allowable sequence at the beginning of a word (***bl**ue*), while **bn** is not. Finally, we also know how sound patterns alternate. For example, in the regular plural formation in English, what is written as ***s*** or ***es*** is pronounced [s], [z], or [ɪz] depending on certain properties of the last sound of the word. As native speakers, without thinking we produce the expected forms (*block[s], herd[z], mattress[ɪz]*). It is this knowledge about sound structure – which sounds occur, what their distribution is, how they can be combined, and how they might be realized differently in different positions in a word or phrase, that constitutes the study of **phonology**.

Central to research in phonology is documenting and characterizing the full range of attested sound structures and patterns across the languages of the world.[1] In this chapter, we explore some of the central generalizations about sounds, using theories and tools that allow us to insightfully analyze these patterns. We will focus on three areas: sound inventories and contrasts (Section 2), structure above the level of the sound unit or segment, that is prosodic organization (Section 3), and structure internal to the segment (Section 4). The general approach followed here is **generative phonology** (see Chomsky and Halle 1968, also Kenstowicz and Kisseberth 1979) where the goal is to develop a theory that accurately models a speaker's knowledge of his or her language. In Section 5, we consider phonology in a broader context, considering alternative views and identifying emerging trends.

2 Inventories and Contrasts

2.1 Inventories

All languages have consonants and vowels. Consonants are sounds with a constriction in the vocal tract, while vowels lack such a constriction. Vowels can serve as the core of a syllable (see below in Section 3), while consonants generally cannot. Consonants must co-occur with vowels to produce forms that are pronounceable. Both consonants and vowels can be defined in terms of where in the mouth and how they are produced. Consonants are characterized in terms of **place** and **manner** of articulation. Place of articulation (e.g., labial, coronal, velar) indicates where the obstruction occurs created by the movement of an "active" (e.g., the tongue) and "passive" (e.g., the soft palate) articulator. The manner of articulation indicates the degree of constriction: complete closure (stops), noticeable obstruction (fricatives) or a combination of closure and obstruction (affricates), closure in the mouth with air escaping through the nose (nasals), or only approximation (liquids and glides). Vowels are generally characterized in terms of the height of the tongue or jaw (high, mid, low) and the relative backness of the tongue (front, central, back) and whether the lips are rounded or unrounded. Other properties play a role, such as whether the vocal folds are close together and vibrating (voiced) or farther apart, allowing freer passage of air from the lungs (voiceless).[2]

So far we have presented examples using English spelling, with some additional pronunciation information provided in square brackets [ ]. English spelling is sorely inadequate for accurately describing the sounds currently used in English. The 26 symbols of the Roman alphabet are not sufficient to represent all of the consonant and vowel sounds of English (as we'll see there are about 39, depending on the dialect). But this isn't the only issue. In order to describe sounds reliably, we need a fully systematic relationship between sound and symbol, something that English spelling doesn't provide, since there are many-to-many correspondences of sound-to-symbol. For example, the sound [k] corresponds to several different symbols or symbol combinations – ***c**at, **k**ite, **kh**an, **q**uite* (**qu** = [kw]), *e**ch**o, pa**ck**, bo**x*** (**x** = [ks]), whereas the letter **c** represents various sounds – [k]: ***c**at*, [s]: ***c**ite*, [tʃ]: ***c**ello* (not including two-symbol combinations, such as **ch**).

Additionally, we often need to be able to include more pronunciation detail. This need for greater detail is true even of languages with much more transparent spelling systems than English. We need what is called **phonetic transcription**. The International Phonetic Alphabet (IPA) is a system of phonetic transcription that allows us to systematically represent the sounds of any language. This system, developed by the International Phonetic Association (founded in 1886) is periodically updated, to reflect changes in thinking on transcription and to include new speech sounds that have been "discovered."

A sound inventory is the set of sounds occurring in a particular language. Looking across the inventories of the languages of the world, we find that the number of consonants and vowels, as well as the specific selection of sounds, varies greatly from one language to another. In his study of the sound inventories of 317 languages, Maddieson (1984, updated Maddieson and Precoda 1990) found that the number of consonants in a language ranged from 6 to 95, with a mean of 22.8; while the number of vowels ranged from 3 to 46 with a mean of 8.7.

Considering this range of sound inventory size, let's see how the inventory of American English compares, as shown in (2). For the consonants (C), the places of articulation are the column headings and the manners of articulation are the row headings. When two sounds appear within a single cell, the one on the left is voiceless (without vocal fold vibration) and the one on the right is voiced (with vocal fold vibration). For the vowels (V), in addition to tongue backness (marking the columns) and height (marking the rows), adjacent pairs within a category differ in "tenseness" (longer and more peripheral in the vowel space, e.g., [i]) vs. "laxness" (shorter and more centralized, e.g., [ɪ]). There are also three diphthongs (vowel-glide combinations that function as a single vocalic unit).

(2) Sound inventory of English

Cs	LABIAL	DENTAL	ALVEOLAR	PALATO-ALV.	PALATAL	VELAR	GLOTTAL
stop	p b		t d			k g	
fricative	f v	θ ð	s z	ʃ ʒ			h
affricate				tʃ dʒ			
nasal	m		n			ŋ	
liquid			r l				
glide					j	w	

Vs	FRONT	CENTRAL	BACK
high	i ɪ		ʊ u
mid	e ɛ	ɝ ʌ	ɔ o
low	æ	a	

Diphthongs: aj, aw, oj

There is some variation in the number of sounds argued to occur in English. This is due in part to dialect differences for example, [a] *co**t*** vs. [ɔ] *ca**u**ght* contrast in some dialects but not others. There are also analytic questions, such as whether the affricates, [tʃ] (***ch**urch*) and [dʒ] (*ju**dg**e*), should be treated as single units or as sequences of sounds. Nevertheless, the characterization of American English in (2) with 24 consonants, 12 vowels and 3 diphthongs is fairly common. Thus English has an average-sized consonant inventory, though notable in its rich array of fricatives. There are whole classes of consonants that English doesn't exemplify, such as clicks, found in languages of Southern Africa. With 12 vowels, English has a relatively rich vowel inventory, especially considering that the distinctions are all made using only the two dimensions of tongue height and backness. (In the inventory above, we haven't included schwa [ə] or r-colored [ɚ] which occurs only in unstressed position.)

We can compare the English inventory with a language that has a rich consonant inventory such as that found in Arabic. In Modern Literary Arabic, we find a very small vowel inventory, only three distinct vowel qualities (though length differences also result in differences in meaning, e.g., [d**u**r] 'turn!' vs. [d**u:**r] 'houses'), but a very rich consonant inventory. Not only are most of the consonants seen in English found here, but there are additional places of articulation, notably at the back of the mouth (uvular and pharyngeal). In addition, there is a contrast between plain consonants and those with a superimposition of a back tongue position (pharyngealization) and finally consonants also contrast for length ([ba**r**a] 'sharpen' vs. [ba**rr**a] 'acquit'). Including all these contrasting dimensions, there are 48 consonants in this variety of Arabic.

While the inventories of English and Arabic might suggest that there is a tendency for languages with large consonant inventories to have correspondingly small vowel inventories and vice versa, this is not necessarily the case. Consider, for example, Rotokas, spoken in Papua New Guinea (the smallest inventory in Maddieson's 1984 database), with a very common five-vowel inventory [i, e, a, o, u], but only six consonants [p, t, k, g, β, ɾ], for a total of only 11 segments.

While there is great variation in the segments that occur in particular languages – Maddieson and Precoda (1990) identify roughly 900 – strong predictions can nevertheless be made about which sounds will occur. Some sounds and categories of sounds are just more common than others. For example, all languages have stops, but not all languages have fricatives. Beyond these basic observations, in many cases the presence of one property implies the presence of something else in the same system; such generalizations are called implicational language universals. For example, if a language has the mid vowels [e, o] (as in English, *bait* [bet] and *boat* [bot]), it is predicted that it will also have the high vowels [i, u] (English *beat* [bit] and *boot* [but]) and the low vowel [a] (English *pot* [pat]); but the converse doesn't hold, as we've seen in Arabic with [i, u, a], but lacking [e, o]. See Hyman (2008) for a discussion of such phonological universals. An ongoing area of investigation is identifying principles (both phonological and phonetic) that play a role in defining attested as well as potentially well-formed sound inventories, such as the principle of "economy" (Clements 2003), whereby contrasting phonological classes tend to be used maximally.

2.2 Contrast

When we characterize the inventory of sounds of a language, we draw an important distinction between those sounds that can be used to make meaningful contrasts in a language vs. those that occur only in predictable contexts. The above discussion of the inventories of English, Arabic, and Rotokas reflects those sounds argued to be **distinctive** or **in contrast** in the language (though, as we discuss below in Section 4, the status of [ŋ] in English is debatable).

In order to determine the status of a sound, we use a simple test for **minimal pairs**. Minimal pairs (or sets) are words with distinct meanings differing only in one sound. Thus we can show that [m] and [n] (differing only in place of articulation) are distinct sounds in English, since the substitution of these sounds alone is enough to change the meaning of a word: ***m**eat* vs. ***n**eat, si**mm**er* vs. *si**nn**er, ra**m*** vs. *ra**n***. The presence of [m] vs. [n] at the beginning, middle, or end of a word results in words with distinct meanings.

If a sound is used distinctively in a particular language, it is a **phoneme** in that language (and is transcribed in slanted brackets / /). Phonemes are argued to be the units encoded in lexical entries (the forms in our mental dictionaries), and upon which speakers judge "sameness" and "differentness." However, phonemes can vary in their actual pronunciation, depending on the context of the neighboring sounds, the structure of the utterance, and so forth.

Two languages may have the same sounds or **phones** (the actual phonetic events, transcribed in [ ]s), but their grouping into phonemes might be different. In English, the sounds [b, p, p^h] all occur (that is, voiced, voiceless unaspirated, and voiceless aspirated). While [p^h] and [b] contrast, the selection of [p] or [p^h] is determined by the phonological context, as schematized in (3a). *Buy* [baj] contrasts with *pie* [p^haj], but the realization of a voiceless stop as aspirated word initially (*pie* [p^haj]) or unaspirated following [s] (*spy* [spaj]) is predictable and there are no minimal pairs for [p] and [p^h]. These three phonetic categories are mapped to only two abstract phonological categories. Yet in Thai, all three sounds occur and can produce differences in meaning, as shown by the minimal set in (3b).

(3) a. English

	phonemes		phones		b. Thai phonemes		
buy [baj]	/b/	———	**[b]**	———	/b/	[baa]	'crazy'
spy [spaj]	/p/	———	**[p]**	———	/p/	[paa]	'aunt'
pie [p^haj]		———	**[p^h]**	———	/p^h/	[p^haa]	'cloth'

but no *[paj] or *[sp^{h}aj]

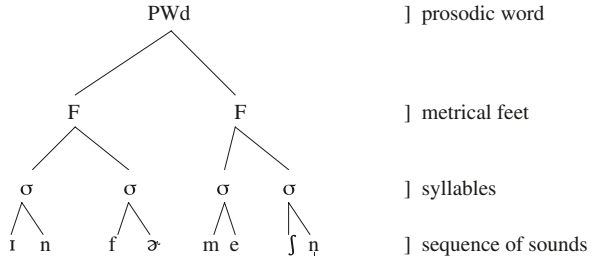
To summarize, these three phones [b, p, p^h] constitute three separate phonemes in Thai, but only two in English.

In English, [p, p^h] stand in a special relationship to each other, since they are part of the same phoneme (usually taken to be /p/). Such sounds are called **allophones**. We can capture this relationship by describing the distribution, e.g., [p^h] occurs at the beginning of words and [p] occurs after [s]. (There is more to this pattern, but we won't pursue it here.) Or we can go a step further and argue that the phoneme /p/ occurs at an abstract or **underlying** level and account for the observed surface distribution with a phonological rule. We return to the issue of rules in Section 4.

3 Structure Above the Level of the Segment: Prosodic Organization

The sound structure of a word (a unit that can be defined on several linguistic levels, including morphologically and prosodically) includes not only the sequence of sounds (made up in turn of bundles of distinctive features, as discussed in Section 4), but also entails the hierarchical grouping of these sounds. Let's take the English word *information* as an example:

(4)



This word consists of a sequence of sounds **ɪ-n-f-ɚ-m-e-ʃ-n̩**. Most speakers of English would agree that this form consists of four syllables (σ) broken up as **ɪn-fɚ-me-ʃn̩**. Consonants and vowels are grouped into syllables in nonarbitrary ways, with a vowel forming the core or **nucleus** optionally flanked by consonants. In the final syllable **[ʃn̩]**, the nucleus is **n̩**, which is a syllabic nasal, functioning as a vowel. These syllables are in turn organized into stress groupings (**ɪ̀n-fɚ**)(**mé-ʃn̩**). The third syllable is the most prominent (primary stress, indicated with an acute accent ´) and the first also has some prominence (secondary stress, indicated with a grave accent `). These patterns of prominence can be accounted for by grouping the syllables together into units known as metrical feet (F). (For an introduction to metrical theory, see Hayes 1995). Finally the feet are grouped together into a Prosodic Word (PWd). The Prosodic Word often has the same shape as a morphologically defined word, but not necessarily. There are, for example, so-called function words (grammatical words), which we take to be words morphologically, but that can't stand on their own phonologically, such as *a*, or *the*. The syllables, feet, and prosodic words together constitute the prosodic structure of a word. Words in turn can be grouped into higher levels of prosodic structure as well, at the phrase and utterance level. This grouping and hierarchical organization together constitute the prosodic organization.

The structure of segments, how segments are combined, and how syllables, metrical feet, and prosodic words are organized, are integral parts of phonology. In this section, we examine syllable structure as an example of prosodic organization above the level of the segment.

3.1 Syllable structure

Syllable structure influences the ways segments are organized, sometimes leading to the insertion or deletion of a segment. Consider an example from Korean, shown in (5), where we observe that sometimes a cluster of consonants occurs and sometimes one of the members of the cluster is deleted. This is an example of what we call an **alternation** where the same morpheme varies in its realization, conditioned by some aspect of the sound system (in this case the allowable syllable structure).

(5) Consonant ~ zero alternations in Korean clusters

root		**+ vowel-initial suffix**		**+ consonant-initial suffix**	
		-a nominalizing suffix		**-t'a** infinitive	
/palp/	'tread on'	palp + a	'treading on'	pap + t'a	'to tread on'
/salm/	'boil'	salm + a	'boiling'	sam + t'a	'to boil'

The basic idea is that in some cases, Korean syllable structure can't accommodate all the segments in the underlying representation, so one is deleted in the surface form. The underlying clusters (/lp/ and /lm/) are allowed to surface before a vowel-initial suffix, since the second member of the cluster can be syllabified as the beginning of the second syllable, producing [palpa] and [salma]. But when the root occurs before a consonant-initial suffix, the first consonant of the cluster (here /l/), is deleted, producing [papt'a] and [samt'a]. (In other cases, the second consonant is deleted instead.) The syllabification of forms with vowel-initial and consonant-initial suffixes respectively is shown in (6) for /palp/ (where < > indicates a segment not incorporated into the syllabic structure):

(6)

This deletion is directly driven by the allowable syllable structure.

As noted in Section 1, restrictions also exist on possible sequences of sounds. For example, in English *[bn] can't occur at the beginning (7a) or end (7b) of a word.

(7) a. *bnick
b. *kibn
c. lab-network
d. drabness
e. Abner

It is not the case that [bn] is always bad in English. In (7c), this sequence is fine, however, the word is a compound and we might argue that it consists of two prosodic words grouped together ([[lab]$_{pwd}$[network]$_{pwd}$]$_{pwd}$) and therefore it is not held to the same restrictions. The fact that (7d) is allowable might be attributed to the sounds belonging to different morphemes (*drab* and *-ness*). But in (7e) there aren't distinct words or morphemes. What then is the difference between [bn] in (7a & b) and (7e)? In the latter case, the [b] and [n] are in different syllables, while in the former they are in the same syllable. The restriction holds of a sequence within a syllable and seems to be due to the fact that [b] and [n] are too similar in terms of **sonority**. Sonority can be defined loosely as the degree of constriction in the mouth during the production of a particular sound. Most important for our purposes is the observation that there is a hierarchy of how **sonorous** sounds are. Vowels are more sonorous than consonants; and within the consonants, further divisions can be made. Nasals, liquids, and glides, together known as **sonorants**, are more sonorous than both fricatives, with constriction creating frication or noise, and stops, where there is complete closure. Stops and fricatives (together with affricates) are known as **obstruents**, since there is a significant obstruction. Thus we find the following widely observed crosslinguistic pattern:

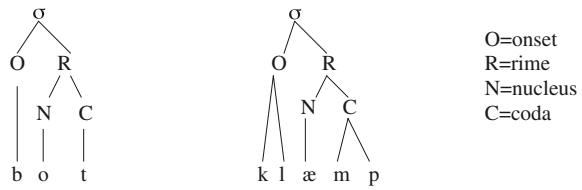
(8) Sonority hierarchy:

more sonorous				**less sonorous**
vowels	>	sonorants	>	obstruents

The sonority hierarchy characterizes the behavior of sounds in syllable structure and many other aspects of phonological patterning. Whether finer-grained distinctions of the sonority hierarchy are required is a question open to much debate. (See Zec 2007 for recent discussion of this question and for an overview of issues in syllable structure more generally.)

Strong evidence exists for making reference to the syllable as part of the hierarchical structure of the phonological system to account for observed alternations and also to capture consonant sequencing restrictions. In addition, the syllable is argued to be divided into subparts. Evidence for this comes from the fact that co-occurrence restrictions hold on the consonants preceding, as well as following, the core of a syllable, but not generally across the subparts of the syllable. One approach to the internal organization of the syllable is shown in (9), where the substructures of *boat* and *clamp* are illustrated:

(9)



Based on a wide range of evidence, there is argued to be a division in the syllable between the **onset** and the **rime** constituents. The division into onset and rime allows us to capture various consonant sequencing restrictions and is also relevant for other aspects of the phonology, as well as language games and poetry. The rime corresponds to the unit that rhymes, e.g., ***oat***, *b****oat***,

*bl**oat***; and the onset is the unit shared in poetic patterns of alliteration, e.g., ***bl**ue, **bl**ow, **bl**ithe, **bl**oat*. Following some views, the rime is further divided into the **nucleus**, the core of the syllable that contains the vowel or vocalic element(s), and the **coda**, which contains any following consonant(s). In English, the only required element of the syllable is the nucleus (e.g., ***oh*** [o], ***I*** [aj]), although in many languages the onset is also an obligatory part of the syllable.

Let's consider some of the co-occurrence restrictions within onsets and codas in English. All of the examples in (10) are well-formed English syllables (and in these cases independent words too).

(10) Examples of possible syllables in English

		coda:							
		Ø		C		CC		CCC	
onset:	Ø	oh	[o]	ode	[od]	old amp	[old] [æmp]	amps	[æmps]
	C	bow	[bo]	boat	[bot]	bolt	[bolt]	bolts	[bolts]
	CC	blow	[blo]	bloat clam	[blot] [klæm]	clamp	[klæmp]	clamps	[klæmps]
	CCC	spree	[spri]	split	[splɪt]	splint	[splɪnt]	splints	[splɪnts]

In English, anything from no consonants to up to three consonants preceding and four following a vowel may constitute a well-formed syllable.[3] Many restrictions hold, however, on possible combinations of consonants preceding or following the vowel and only a small subset of logically possible combinations occurs. For example, in English triconsonantal onsets ($C_1C_2C_3$), the first sound (C_1) must be [s], followed by a voiceless stop ([p, t, k]), followed by a liquid ([r, l]) or glide ([j, w]). Many of the occurring patterns can be characterized with reference to the sonority hierarchy (8), though other factors also come into play. Thus in CCC onset clusters the pattern of C_2 and C_3 follows the sonority hierarchy, with onsets showing a rise in sonority going from C_2 to C_3: stops followed by the more sonorous liquids and glides. But the occurrence of [s] preceding such clusters is not predicted even with modification to the sonority hierarchy, since [s] is not less sonorous than the stops, and therefore requires a distinct explanation. Similarly in characterizing what coda clusters can occur in English, sonority also plays an important role. In general, the first member of a two member coda cluster must be of the same or greater sonority than the second member (e.g. *le**nt**, be**lt**, li**ft**, mi**st**, a**pt***). Such patterns can be characterized straightforwardly, making reference to the subparts of the syllable, but are much harder to characterize if we only refer to the string of segments.

How much explicit or formal internal structure to the syllable is warranted and how it should be encoded are much debated questions; however, reference to some degree of substructure of the syllable allows for insightful generalizations about sequencing restrictions and other aspects of sound distribution. Indeed in English, we can capture the pattern presented in (7) by observing that the sequence [bn] cannot occur together as part of an onset or coda. In addition, reference to syllable subconstituency enables us to capture the broader distribution of sounds in many cases. For example, as noted in (1), the distribution of /h/ in English is limited: it can occur only in the onset of a syllable (and if it is not word-initial, only if the syllable is stressed, e.g., *vehicle* [víɪkl̩] vs. *vehicular* [vi**h**íkjələ˞]).

While it is relatively straightforward to count the number of syllables in a word, it is often trickier to decide where the syllable break falls. Crosslinguistically in forms of the shape (C)VCV,

the division generally falls before the medial C, (C)V$CV ($ indicates a syllable break). However, in English, the situation is additionally complicated by the stress pattern. In words such as those in (11a), it is widely agreed that the syllable divisions are as shown, characteristic of the strong crosslinguistic tendency.

(11) a. attáck [ə$tǽk]
belów [bə$ló]

b. áttic [ǽtɪk]
béllow [bɛ́lo]

However, many researchers have argued that in the cases such as (11b), the medial consonant belongs either to the first syllable or is shared by the two syllables in order to account for otherwise systematic observations about the relationship between syllable structure and stress in English. (Even though the middle consonants in some of the forms in (11) are written with a doubled letter (tt, ll), they are just single consonants.)

In English, in the case of (C)VCCV(C), the syllabification depends on the specific sequence of consonants. If the CC is an allowable word onset (and therefore an allowable syllable onset), the syllable division is before both consonants (12a), but otherwise it is between the two consonants (12b).

(12) a. apply [ə$**pl**aj] cf. plea [pli]
abrupt [ə$**br**ʌpt] cf. brush [brʌʃ]

b. Adler [æ**d**$**l**ɚ] * [dli]
Abner [æ**b**$**n**ɚ] * [bni]
ardent [a**r**$**d**ṇt] * [rdi]

Other languages show much greater restrictions on syllable structure than English does. Consider some examples from Japanese in (13).

(13) Allowable syllables in Japanese: CV, V, CVN, CVC

a. CV, V

[ki]	‘tree’
[kokoro]	‘heart’
[mado]	‘window’
[tegami]	‘letter’
[ito]	‘string’
[origami]	‘paper folding’

b. N$C

[tombo]	‘dragonfly’
[hantai]	‘opposite’
[neŋkin]	‘pension’

c. C$C

[kitte]	‘stamp’
[onna]	‘woman’
[hakka]	‘peppermint’
[kaʃʃa]	‘pulley’

As illustrated in (13), only (C)V(:) and (C)V(:)C syllables occur in Japanese. CV syllables can occur in any position in the word (13a). But CVCs are allowed only if the coda consonant is a nasal (13b), or part of a geminate (long consonant) (13c), and in these cases usually followed by another syllable. Thus [tom] is a well-formed syllable when followed by [bo], but it would not be allowed if it occurred on its own or as the final syllable in a word. A final alveolar nasal (as in [neŋki**n**] above in (13b)) is well-formed, but other nasals and other consonants in word-final position are not allowed.

Additional evidence for the allowable patterns can be seen by looking at the ways foreign words are modified when they are borrowed into Japanese. Let's consider what happens to some words borrowed from English, as shown in (14).

(14) Borrowings from English into Japanese:

	word	**English**	**Japanese**
a.	pin	[pɪn]	[pin]
	pie	[paj]	[paj]
	Chicago	[ʃɪkago]	[ʃikago]
b.	million	[mɪljən]	[mirion]
	avocado	[avəkado]	[abokado]
	rally	[ræli]	[rari:]

Some words are borrowed basically as is (14a), or with modifications to any nonoccurring segments, with these being substituted by a similar sound that does occur in Japanese (14b). (The lax vowels of English [ɪ, ɛ, ʊ, ɔ] are realized as short [i, e, u, o] and the tense vowels [i, e, u, o]) are realized as being long [i:, e:, u:, o:] in Japanese.)

Of particular interest are cases where nonallowable consonant clusters occur; in such cases, Japanese uses the strategy of adding extra vowels (epenthesis), as illustrated in (15):

(15) More borrowings from English into Japanese:

	word	**English**	**Japanese**
a.	free	[fri]	[fUri:]
	spray	[spre]	[sUpUre:]
b.	peak	[pik]	[pi:kU]
	kiss	[kɪs]	[kisU]
	Bill	[bɪl]	[birU]
	beat	[bit]	[bi:tO]
c.	speed	[spid]	[sUpi:dO]
	cross	[krɔs]	[kUrosU]
	test	[tɛst]	[tesUtO]
	street	[strit]	[sUtOri:tO]
	contrast	[kantræst]	[kontOrasUtO]
	baseball	[besbɔl]	[basUbarU]

Consider first cases with onset clusters in (15a). The inserted vowels are indicated in upper case symbols. (The vowel that is inserted in these cases is usually [u] (U), except after alveolar stops, where an [o] (O) is inserted.) In (15b) we see borrowings of monosyllables of the shape CVC.

Here final vowel epenthesis occurs since a consonant can occur in coda position only if it is followed by an appropriate consonant in the next syllable in the same word. Finally cases with both an onset cluster and final consonant or consonant cluster are shown in (15c). All of these clusters are broken up into many more syllables in Japanese than found in the original English source, with the exception of [nt] in *contrast* which is well-formed in Japanese (cf. 13b).

In the case of nonallowable clusters in borrowed words, some other languages delete segments. Consider what happens to final consonant clusters in Indonesian in words borrowed from English (or Dutch). In Indonesian, the allowable syllable structure is (C)V(C), so final clusters in borrowed words pose a problem. As shown in (16), the final clusters are simplified by deleting the final consonant (similar to the pattern seen for Korean above in (5), although in those examples, the first member of the cluster was deleted).

(16)

word	**English**	**Indonesian**
sport	[spɔrt]	spor
aqueduct	[ækwədəkt]	akuaduk
tolerant	[talə˞rn̩t]	toleran
test	[tɛst]	tes

To account for such systematic syllable patterns, phonologists have proposed various devices including rules, templates, well-formedness conditions, and constraints.

3.2 A constraint-based account

A current approach, **Optimality Theory**, involves the idea of competing phonological constraints, which can be ranked in importance with respect to one another. Due to such ranking, a less important constraint can be violated in order to obey a more important constraint that it conflicts with. Languages differ in how they rank particular constraints. If we have correctly identified the relevant constraints (a major research agenda in itself), then the set of logically possible rankings of those constraints should match up with the range of sound patterns seen across languages. (See Kager 1999 and McCarthy 2008 for an introduction.) Optimality Theory offers an insightful account of syllable patterns and makes strong predictions about allowable syllable types crosslinguistically, and it also accounts for certain implicational universals such as the fact that if a language allows CVC syllables it will also allow CV syllables and if it allows V syllables, again it will also allow CV ones.

As widely discussed, the ideal syllable is CV. Syllables minimally consist of a vowel; onsets are preferred; and codas are dispreferred. To account for the preference for CV syllables as well as the range of crosslinguistic variation observed in syllable structure, two general sorts of constraints interact. First there are markedness constraints – constraints that capture systematic crosslinguistic biases. The preference for CV syllables has been argued to emerge from three markedness constraints, stated here informally:

(17) Syllable structure markedness constraints:

	constraint	**informal definition**
a.	Nuc	Syllables must have a nucleus
b.	Onset	Syllables must have an onset
c.	NoCoda	Codas are not allowed

If this were the whole story, all languages would have only CV syllables, but this is clearly not the case (look at English!). There are also constraints that mediate between the underlying representation (the input to the constraints) and the actual realization of the form (the output of the

constraints). The two constraints relevant for our purposes, again stated informally, limit how different the input and output can be. (* = Don't)

(18) Input/output constraints

	constraint	informal definition
a.	*Add	Only the material of the input should appear in the output; don't add material to the input [standardly called DepIO]
b.	*Delete	Underlying material should be incorporated in the output; don't delete material from the input [standardly called MaxIO]

Other constraints can also affect syllable structure, but these five constraints suffice for our discussion. To test constraint rankings, we compare the input of a form and a list of possible (expected) outputs (placed in the leftmost column of what is termed a *tableau*) with respect to a particular ranking of the relevant constraints (placed in columns, going from higher to lower ranking as we go from left to right). No matter what the relative ranking of these five constraints in a particular language, if we have an input or underlying form of the shape CV(CV)(CV), then all of the above constraints – those affecting syllable structure and those affecting input/output relations – can be satisfied. This is true in both English and Japanese, as shown in (19a) for English *banana* [bənænə] and (19b) for Japanese [kokoro] 'heart'. In these tableaux, the constraints are all unranked, in contrast to the tableaux that follow, where they are ranked.

(19) a. English *banana* [bənænə]

/bənænə/	Nuc	Onset	NoCoda	*Add	*Delete
[bə$næ$nə]	√, √, √	√, √, √	√, √, √	√, √, √	√, √, √

b. Japanese [kokoro] 'heart'

/kokoro/	Nuc	Onset	NoCoda	*Add	*Delete
[kokoro]	√, √, √	√, √, √	√, √, √	√, √, √	√, √, √

A checkmark in the relevant cell indicates that the constraint is met; there are three checkmarks in each cell referring to each of the three syllables in these cases. Onset and NoCoda together (no matter what their ranking) ensure that an intervocalic consonant (VCV) will be syllabified as the onset of the second syllable (V$CV).

Let's now consider some cases where the same input results in different outputs in different languages. Consider the English word *test*, which as we saw above is realized as [tesuto] in Japanese and [tes] in Indonesian. (I leave Nuc and Onset out of the following discussion, as they are met by all of the cases we are considering.)

In English, the input [tɛst] matches the output, even though it violates NoCoda twice. This provides evidence that NoCoda is ranked below both *Add and *Delete. In other words, in English meeting the requirements of the input/output constraints is more important than adhering to the markedness constraints. We show this by comparing possible outputs (20a–e) and showing that with a given ranking, the optimal or best-formed candidate (20a, [tɛst], indicated by ☛) is the observed output, while the others (20b–e) fail to surface.

(20) English *test* [tɛst]

/tɛst/		*Add	*Delete	NoCoda
a. ☛	[tɛst]	√	√	**
b.	[tɛs]	√	*!	*
c.	[tɛ]	√	*!*	√
d.	[tɛstV]	*!	√	*
e.	[tɛsVtV]	*!*	√	√

When a form violates a constraint with respect to the input, we mark that violation with "*". The optimal candidate violates the highest-ranked (leftmost) constraint the least number of times, if at all. In case of a tie at that level, lower-ranked constraints also come into play. An ! indicates an insurmountable violation. This is followed by shading of the successive cells in the same row, indicating that the adherence to these lower-ranked constraints isn't relevant to the outcome. (20a) is the optimal candidate in this case, even though this form violates NoCoda twice. This is still preferable to a violation of either *Add (20d & e) or *Delete (20b & c), providing evidence that both of these constraints outrank NoCoda (hence NoCoda is positioned to the right, separated by a solid vertical line). Since both *Add and *Delete have to be met, we don't have evidence from this example for their relative ranking in English.

The pattern in Japanese is very different. In Japanese, priority is given to the markedness constraints over the input/output constraints. In order to meet the high ranking NoCoda constraint, vowels are inserted, providing evidence that *Delete outranks *Add, as shown in the tableau in (21):

(21) Japanese [tesuto] 'test'

/tɛst/		NoCoda	*Delete	*Add
a.	[test]	*!*	√	√
b.	[tes]	*!	*	√
c.	[te]	√	*!*	√
d.	[testV]	*!	√	√
e. ☛	[tesVtV]	√	√	**

We see here that (21e) [tesuto], which respects both NoCoda and *Delete, is the optimal candidate. We use V to represent an inserted vowel. The actual quality of the inserted vowel is assumed to be a language-specific question. As we saw above in (13b & c), Japanese tolerates some limited violations of NoCoda. However, these codas cannot have their own place specification, rather they must share it with the following onset consonant, either as part of a geminate or as part of a nasal-stop cluster agreeing in place of articulation.

Finally in Indonesian, we find a case where deletion is tolerated, indicating the relatively low ranking of *Delete. This is balanced with a violation of NoCoda, since the optimal form involves

one violation of each NoCoda and *Delete (in contrast to English that violates NoCoda twice and Japanese that violates *Add twice).

(22) Indonesian [tes] 'test'

/tɛst/		*Add	NoCoda	*Delete
a.	[test]	√	*!*	√
b. ☛	[tes]	√	*	*
c.	[te]	√	√	*!*
d.	[testV]	*!	*	√
e.	[tesVtV]	*!*	√	√

The optimal candidate in Indonesian is (22b). Our analysis accounts for the fact that both (22d & 22e) are eliminated, but more needs to be said about why the optimal outcome is (22b) rather than (22a or 22c). An additional constraint must be involved; the intuition is that a single consonant in coda position is more acceptable than a cluster and also, there is a limit to how much deletion the system will tolerate. There is also more to the story, since in the case of onset clusters, vowels are inserted rather than consonants being deleted, for example, Indonesian *stasiun* [sətasiun] from Dutch *station*, but we leave aside these additional details in our current discussion.

There are clearly additional complexities, since all three languages allow vowel-initial words (hence limited violations of Onset) and more needs to be said about why, in Japanese, a final syllable such as [kin] is allowed but one such as [tom] is not. Finally, additional constraints are needed to account for the division of medial consonant clusters into codas and onsets, e.g., English *abrupt* [ə$brʌpt] vs. *Abner* [æb$nɚ]. In many languages, VCCV will surface as V$CCV if CC is an allowable onset (again additional constraints are required). If CC is not an allowable onset, the VC$CV syllabification would be the optimal candidate.

While I haven't provided a complete account of these three cases, we can see that the relative ranking of this limited set of constraints allows us to capture these different strategies of syllabification. Other languages are predicted to show different outputs. For example, the output form [testV] would be optimal in a language that had some tolerance of single consonant codas (like Indonesian) but which ranked *Delete over *Add.

In this section we have seen that grouping of sounds into syllables and subsyllabic constituents offers a more insightful account of sound patterns than one where only reference to the segment is made. In addition we have looked briefly at how a constraint-based approach, with minimal constraint violation, allows us to account for some of the crosslinguistic variation observed in syllable structure. Grouping of sounds into syllables is the lowest level of prosodic organization.

4 Subsegmental Structure

4.1 *Features and segmenthood*

Till this point, we have focused on segments and larger units. Good evidence for the psychological reality of segments exists, including speaker intuition, alphabetical writing systems, speech errors, and the fact that phonological alternations involve such units. But there is also good evidence that segments are made up of smaller units and that by making reference to these smaller

units, a more insightful discussion of sound patterning is possible. We have an intuition that [p, b] are more similar to each other than [p, l] are. This is because the former share more sound properties than the latter. These sound properties are called **distinctive features**. The notion of distinctive features grows out of the work of Trubetzkoy, Jakobson, and others (see Anderson 1985 for an excellent survey of the history of phonology). While numerous specific systems have been proposed, most current systems have developed from Chomsky and Halle (1968). Most approaches to phonology assume some kind of feature system and take the features to be the smallest building blocks of phonology. Segments thus consist of bundles of features, or feature matrices, as exemplified for *bill* /bɪl/, in (23):

(23) feature matrices

b	ɪ	l
+consonantal	-consonantal	+consonantal
-continuant	+high	-continuant
-sonorant	-back	+sonorant
-nasal	-tense	-nasal
labial		coronal
		lateral

Evidence for specific feature proposals comes from their adequacy in capturing the recurrent crosslinguistic grouping of sounds, referred to as **natural classes**. Take, for example, the feature [±sonorant] (where [+sonorant] defines the class of sonorants and [-sonorant] defines the class of obstruents). The feature [+sonorant] is defined as that class of sounds for which spontaneous vocal fold vibration (or voicing) is possible, including nasals, liquids, glides, vowels, sounds for which there is not a close obstruction of the vocal tract. In the typical case, sonorants are voiced and do not show a contrast between voiced and voiceless. For obstruents ([-sonorant]: stops, fricatives, and affricates), voicing involves certain articulatory adjustments to maintain subglottal air pressure and keep the vocal folds vibrating. For obstruents, the least marked category is voiceless, but obstruents often show a contrast between [+voice] and [-voice]. A strong implicational universal is that if a language has a voicing contrast among sonorants (as found, for example, in Burmese), then it also has a voicing contrast among obstruents. The natural class defined by [±sonorant] is also shown by syllabic consonants in English (the nasals and liquids in words like *bottle* [baɾl̩] and *button* [bʌʔn̩]) and the division between the sonorants and obstruents is crucial to the sonority hierarchy discussed above.

A striking fact is that features themselves are not arbitrary classificatory elements, but rather are closely linked to phonetic structure. Thus we find a convergence of phonetic events and the sounds that are found to pattern together in the phonologies of language after language. I will not provide a systematic discussion of distinctive features, since a number of good overviews are available (see, for example, Clements and Hume 1995). I refer here rather informally to specific features.

Many interesting and important issues surround the status of features. The fundamental question is whether features are universal, with the same set characterizing the phonological patterns in all languages, and if so why. The dominant view (at least till recently) was that features are universal because they are innate (part of the human language endowment). This has led to much debate about a specific set of features that can account for all the occurring sounds in the languages of the world. If the striking patterns of sounds across languages are *not* due to an innate set of features, the questions to be answered are different, but equally interesting. How are distinctive features learned and to what degree are they similar or identical across languages? Much current work in phonetics, phonology, and language acquisition addresses itself to these questions. (For discussion and review of this literature, see Cohn 2011; Mielke 2008). Here I assume that phonological features are roughly equivalent across languages and leave open the question of how they are acquired.

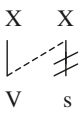
Often the patterning of sounds is characterized in terms of the specific featural content of segments, but other times the presence or absence of segments themselves accounts for an observed pattern. Thus sometimes it is appropriate to refer to the segment as a unit independent of its featural content. In some cases a segment is deleted without leaving any evidence behind (such as the Korean consonant deletion case illustrated above in (5)), but in other cases, the timing of a deleted segment "stays behind." This is called **compensatory lengthening**. Consider the widely discussed case from Latin illustrated in (24).

(24) Latin compensatory lengthening

/ko**s**mis/	[ko:mis]	'courteous'
/ka**s**nus/	[ka:nus]	'gray'
/fide**s**lia/	[fide:lia]	'pot'

We see in (24) that an /s/ is deleted before another consonant (/m, n, l/). But the /s/ doesn't completely disappear; rather it leaves behind its timing unit (indicated here by an X), resulting in a lengthening of the preceding vowel. We can capture this change as follows (where **V** and **s** informally represent the relevant bundles of features).

(25)



The feature bundle of /s/ is deleted, but its timing unit is reassociated with the preceding vowel. Direct reference to the timing aspect of a segment allows us to capture this straightforwardly.[4]

4.2 Alternations

With these further refinements of the representation of phonological units – features organized into segments and timing units, in turn grouped into larger units – we are ready to consider one of the central observations in phonology. Oftentimes phonemes are realized in different ways in different contexts as determined by position in the word, neighboring sounds, (un)stressed position, and so forth. Such differences in the realization of a phoneme are the clearest evidence of the effects of phonology. As seen above, alternations can result from aspects of the higher level organization (for example, in the consonant ~ zero alternations in Korean due to syllable structure). But effects are also found due to the quality of neighboring segments. To take a simple example from English, the prefix /ɪn-/ 'not' changes its shape depending on the following consonant:

(26) Shape of /ɪn-/ prefix in English

[ɪn]
inappropriate
intolerant
indecent

[ɪm]
impossible
imbalance

[ɪŋ]
incoherent
inglorious

The nasal becomes more similar to the following consonant by sharing its place of articulation, with a coronal nasal [n] before coronals (and also vowels), a bilabial nasal [m] before bilabial stops, and a velar nasal [ŋ] before velars. The morpheme /-ɪn/ has several allomorphs including [ɪn-, ɪm-, ɪŋ]. This is an example of **assimilation**, whereby a sound becomes more similar to its neighbor(s). While such patterns of nasal place assimilation are very common crosslinguistically, this pattern is not as systematic in English as in some other languages, since a nasal consonant doesn't always share the place of articulation of the following consonant. For example, in forms compounded with the particle /ɪn-/, for some speakers, assimilation doesn't take place: cf. *input*, [n-p] *income* [n-k]. The (non)application of assimilation in this case is argued to be due to a difference in morphological and prosodic structure between these cases.

It is also assimilation, in this case, of voicing, that accounts (in part) for the alternation in the shape of the regular plural marker in English that we saw above in (1). As observed above, what is spelled as *s* or *es* is pronounced as [s], [z], or [ɪz]. The distribution of these three variant shapes of the plural morpheme is not arbitrary. Rather the distribution is systematically determined by the voicing and place of articulation of the final sound of the stem:

(27) Shape of plural marker in English

a. [s]		b. [z]		c. [ɪz]	
cap	[p]	cab	[b]	match	[tʃ]
cat	[t]	fad	[d]	judge	[dʒ]
book	[k]	dog	[g]	mess	[s]
		can	[n]	buzz	[z]
		file	[l]	wish	[ʃ]
		bow	[o]	garage	[ʒ]

If the final sound of the stem is voiceless, as shown in (27a), then the shape of the plural marker is [s]. (This holds systematically for the stops, but the situation with voiceless fricatives is more complicated: sometimes the voiceless fricative itself becomes voiced and then takes the voiced allomorph [z], such as *leaf* [f], *leaves* [vz], but sometimes the pattern for the stops is found, *chef* [f] *chefs* [fs].) As shown in (27b), if the final sound of the stem (whether an obstruent, sonorant, or vowel) is voiced, then the shape of the plural marker is [z]. Thus the voicing of the final sound in the stem conditions the shape of the plural marker, which agrees in voicing with that sound. But there is a systematic exception to the pattern seen in (27a & 27b), as illustrated in (27c). If the final sound is either an affricate [tʃ, dʒ], or an alveolar or palato-alveolar fricative [s, z, ʃ, ʒ], then the shape of the plural marker is [ɪz]. The intuition here is that [s] or [z] added to stems ending in these sounds would be too similar to be perceptually distinct and so a vowel is inserted to break up the cluster. While some limited exceptions exist, such as *mouse-mice, sheep-sheep, child-children*, there is good evidence that speakers intuitively know the "rule" that accounts for the phonetic shape of the plural marker. Such evidence comes from the fact that both children acquiring English and adults when faced with new words added to the language apply these rules in forming the plural, for example, *macs* [s] and *PCs* [z] and some people even say *mouses* [ɪz].

Such patterns of assimilation are very common across the languages of the world. This is an area where we see a close parallel between the phonology and phonetics. It is a common property of speech that neighboring sounds are coarticulated, that is, that the articulation of adjacent sounds overlaps. Such phonetic effects can become exaggerated and over time result in phonological assimilation. Let's consider another example, the case of vowel nasalization in Sundanese (a regional language of Indonesia).

(28) Sundanese vowel nasalization

a.	[atur]	'arrange'	[ŋãtur]	'arrange' (active)
	[obah]	'change'	[ŋõbah]	'change' (active)
	[parios]	'examine'	[mãrios]	'examine' (active)

b.	[tiis]	'relax in a cool place'	[nĩĩs]	'relax in a cool place' (active)
	[saur]	'say'	[ɲãũr]	'say' (active)

In Sundanese, an initial vowel or one following an oral consonant is oral, while one following a nasal consonant is nasalized. This alternation between nasalized and oral vowels can be seen in corresponding bare stems and active forms, since the active is formed by adding [ŋ] or [+nasal] to the initial consonant of the root, as shown in (28a). Not only is a single vowel following a nasal consonant affected, but a sequence of vowels becomes nasalized, as shown in (28b). Such examples illustrate the importance of distinctive features for an adequate description of such alternations. If we couldn't make reference to a single feature (e.g. [voice] or [nasal]) or set of features (needed, for example, to account for nasal place assimilation), we would be missing a fundamental insight into what is going on in such cases. Within the generative framework, following the seminal work of Chomsky and Halle (1968), *The Sound Pattern of English* (*SPE*), such patterns are accounted for by rules of the form: a → b/ c __ d, "a becomes b in the environment following c and preceding d." The general rule schema offers a formalism for accounting for observed phonological alternations. Rather than just describing the distribution of the differing allophones, this rule formalism incorporates the fundamental idea that one of the variants is basic, or underlying, and that the other variant(s) are derived by rule. Such rules are an attempt to capture the knowledge that a speaker has about the sound patterns of his or her language. Alternatively such patterns can be analyzed as the interaction of violable constraints (as discussed for syllable structure in Section 3.2).

Following the *SPE* rule-based approach, the pattern of nasalization in Sundanese can be represented as shown in (29a), with an example of the application of the rule or "derivation" in (29b).

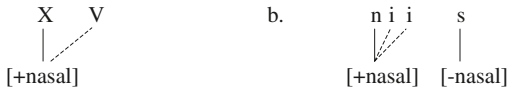
(29) Sundanese Vowel Nasalization:

a. V → [+nasal] / [+nasal] ____
Condition: applies iteratively
"A vowel becomes [+nasal] when it is in the environment following a sound that is [+nasal]"

b.	Underlying representation	/tiis/	/[+nasal] + tiis/
	Vowel Nasalization	—	nĩis
	iterative		nĩĩs
	Surface representation	[tiis]	[nĩĩs]

Further developments suggest that a more accurate account follows from the idea of assimilation as "feature spreading," rather than the changing of feature values. This is part of a more general approach termed **autosegmental phonology**, where specific features can function independently of segments. Following this approach, we could characterize vowel nasalization in Sundanese as follows:

(30)



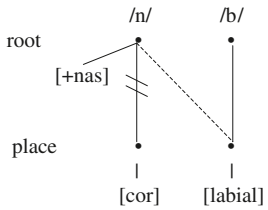
The autosegmental rule in (30a) indicates that the [+nasal] feature specification spreads to the right to a following vowel, resulting in structures such as that illustrated in (30b). Here the pattern of assimilation is captured directly through the sharing of a single feature specification. This has the added advantage of allowing us a straightforward account of the iterative nature of this process.

We also saw an example of spreading in our characterization of compensatory lengthening above in (25), where the whole feature matrix specifying the vowel is shared between the vowel's

timing unit and the following timing unit, freed up by the loss of the feature matrix of the /s/. Viewed in this way, this too can be seen as a sort of assimilation, in this case, total assimilation.

In addition to assimilation of a single feature (e.g., vowel nasalization) and total assimilation (e.g., compensatory lengthening), there are cases where two or more features systematically pattern together. Such is the case of nasal place assimilation, exemplified above for the English prefix /ɪn-/. Cases where a particular set of features pattern together in assimilation and other phonological processes provide strong evidence for the grouping of features (see Clements and Hume 1995 and work cited therein). This general approach, termed **feature geometry**, not only captures the notion of the segment as a unit independent from its featural content (represented by a **root node**), but also offers an explicit proposal of hierarchical structure or subgrouping of features, making direct reference to elements such as the **place node**. An account of nasal place assimilation following this approach is schematized in (31).

(31) Nasal Place Assimilation



Segments can influence each other in a wide variety of ways. There is a rich array of patterns of assimilation, including cases where the segments affecting each other are not adjacent, such as **vowel harmony** where vowels agree in a certain property (e.g., height, backness or rounding), irrespective of the quality of the intervening consonants. For example, suffixes in Turkish vary in their shape depending on the backness of the vowel(s) in the root. Thus *balta* 'axe', *balta**lar*** 'axes', but *kedi* 'cat', *kedi**ler*** 'cats'.

The contrast between segments can be lost in a particular environment. This is known as **neutralization**. Feature changes may be brought about due to the segmental context (that is, influence of neighboring segments), but it is also the case that prosodic structure can drive such effects. It is quite common that the range of contrasts is more restricted in syllable codas than in syllable onsets. One very common pattern of neutralization is what is known as **Final Devoicing**. Consider the following example from Polish:

(32) Polish voicing alternations

a.	klu**p**	'club' sg.		klu**b**i	'club' pl.
	tru**t**	'labor' sg.		tru**d**i	'labor' pl.
b.	tru**p**	'corpse' sg.		tru**p**i	'corpse' pl.
	ko**t**	'cat' sg.		ko**t**i	'cat' pl.

We see an alternation in the voicing of the final consonant of the stem. Just looking at the forms in (32a), we might think that either the voiceless stops are underlying and become voiced between vowels, or that the voiced stops are underlying and become voiceless at the end of the word. But looking at the data in (32b), we see that not all cases show the same alternation; here a voiceless stop surfaces in both forms. This makes it clear that the voiced stops are becoming voiceless. We also note that this pattern seems to apply to a natural class of sounds, in this case, the stop consonants. As shown in (33), this pattern also applies to velar stops (33a), and to fricatives (33b), that is, the class of obstruents (characterized as [-sonorant]). We can capture this pattern by positing underlying forms as shown in (33c) and applying a rule of Final Devoicing,

which can be characterized in *SPE* terms as shown in (33d). Or we can account for such patterns in an autosegmental notation with the delinking of the relevant feature specification, in this case [+voice] (33e). In either case, we can see that the rule works by looking at sample derivations in (33f).

(33) Polish Final Devoicing

a. wu**k** 'lye' sg. wu**g**i 'lye' pl.

b. gru**s** 'rubble' sg. gru**z**i 'rubble' pl.

c. underlying forms:

/klub/	'club'
/trud/	'labor'
/trup/	'corpse'
/kot/	'cat'
/wug/	'lye'
/gruz/	'rubble'
/-Ø/	singular
/-i/	plural

d. [-sonorant] → [-voice] / ____ # (# = word boundary)
"A member of the class of [-sonorant] becomes voiceless in word final position."

e. root [-son] #

larygneal [+voice]

f.

Underlying representation	/klub + Ø/	/klub + i/
Final Devoicing	klup	—
Surface representation	[klup]	[klubi]

We can also capture this pattern in Optimality Theoretic terms, as constraint interaction. There is a tension between a markedness constraint avoiding voiced codas, and an input-output constraint requiring faithfulness to the underlying [+voice] specification. Following Kager (1999: 14–16), the relevant constraints are *Voiced-Coda "Obstruents must not be voiced in coda position," which is in conflict with Ident-IO(voice) "The specification for the feature [voice] of an input segment must be preserved in its output correspondent." In Polish, *Voiced-Coda outranks Ident-IO(voice), resulting in the output form [klup] as illustrated in the tableau in (34).

(34) Polish, /klub/ [klup] 'club' sg.

/klub/		*Voiced-Coda	Ident-IO(voice)
a.	[klub]	*!	
b. ☛	[klup]		*

In English where voicing is maintained in codas, Ident-IO(voice) outranks *Voiced-Coda, resulting in a form faithful to the input voicing specification, as shown in (35).

(35) English, *club* /klʌb/ [klʌb]

/klʌb/		Ident-IO(voice)	*Voiced-Coda
a. ☛	[klʌb]		*
b.	[klʌp]	*!	

Before concluding this section, let's return to the question raised above about the status of [ŋ] in English. While we included /ŋ/ in the chart of the sound inventory in English presented in (3) above, we also noted in (1) that [ŋ] has a defective distribution. One approach to this would be to say that /ŋ/ just has a defective distribution similar to /h/, end of story. Yet this would leave a number of distributional observations unaccounted for. Consider the distributions of the three nasals of English, [m, n, ŋ] in (36):

(36)

	initial	medial	final	N-Vstop	N+V stop	N-Vstop	N+Vstop
m	map	dimmer	dim bomb	camper	amber	camp	—
n	nap	sinner	sin	canter	candor	can't	land
ŋ	—	singer[5] [ŋ]	sing [ŋ]	canker [ŋk]	anger, finger [ŋg]	bank [ŋk]	—

English nasals [m, n] can occur in word-initial position, as well as medially and finally. They can also occur before an oral stop, either medially or finally (except that [mb] doesn't occur as a cluster within a syllable coda, hence *bomb* [bam], but *bombardment* [bəmbardmn̩t]). English nasal [ŋ], on the other hand, doesn't occur in word-initial position. Basically [ŋ] only occurs in the syllable coda, not in the onset. This generalization accounts for its absence word initially and accounts for all the cases except *singer*. Notably *singer* consists of the root *sing* plus the suffix *-er* and so the [ŋ] is, in effect, in syllable-final position until the suffix is added.

This generalization accounts for the distribution, but doesn't explain why it should be so. As noted above, sometimes sounds are limited in their distribution, but crosslinguistically we usually find if a consonant is limited, the greater restriction holds in the coda, not the onset. In other words, neutralization (such as Final Devoicing) tends to occur in codas, not onsets. If we take the spelling as a cue, a solution presents itself. We might argue that [ŋ] is not part of the underlying inventory of English, but rather that it is derived from /ng/ or /nk/ sequences. Briefly the analysis would work as follows. The underlying nasal consonants in English are /m, n/. As noted above, English has a rule of Nasal Place Assimilation whereby a nasal assimilates to a following stop (schematized above in (31)). Based on the evidence from the lack of word-final [mb] clusters, we might also posit a rule of Voiced Stop Deletion that applies to noncoronals, whereby a voiced stop following a nasal consonant is deleted word finally (37a). Given the underlying representations in (37b), the rules of Nasal Place Assimilation and Voiced Stop Deletion together (and an analysis for *singer* that we won't develop here) account for the observed patterns, as shown in the derivations in (37c).

(37) a. Voiced Stop Deletion

$$\begin{bmatrix} \text{-sonorant} \\ \text{-continuant} \\ \text{+voice} \\ \text{-coronal} \end{bmatrix} \rightarrow \emptyset \; / \; \begin{bmatrix} \text{+consonantal} \\ \text{+nasal} \end{bmatrix} \; \underline{\qquad\qquad} \#$$

"A voiced noncoronal stop is deleted word finally following a nasal consonant."

b. Underlying representation:
/dɪm/
/banb/[6]
/bænk/
/sɪng/
/fɪngɚ/

c.

Underlying representation	/banb/	/bænk/	/sɪng/	/fɪngɚ/
Nasal Place Assimilation	bamb	bæŋk	sɪŋg	fɪŋgɚ
Voiced Stop Deletion	bam	—	sɪŋ	—
Surface representation	[bam]	[bæŋk]	[sɪŋ]	[fɪŋgɚ]

When /n/ occurs before a velar consonant it assimilates in place of articulation and if the stop is voiced, it is then deleted. These rules are crucially ordered, as shown by the derivation for *bomb* and *sing* in (38). If Voiced Stop Deletion applies first, it would remove the trigger for Nasal Place Assimilation.

(38) Derivation – wrong order

Underlying representation	/banb/	/sɪng/
Voiced Stop Deletion	ban	sɪn
Nasal Place Assimilation	—	—
Surface representation	*[ban]	*[sɪn]

Thus Nasal Place Assimilation must apply first. The result of Nasal Place Assimilation applying first is that it avoids Voiced Stop Deletion taking away or "bleeding" inputs to Nasal Place Assimilation. This particular type of rule ordering is called "counter-bleeding."

Some additional evidence for this analysis of [ŋ] comes from spoonerisms (speech errors where segments are transposed). In a corpus collected by Fromkin (1971), the following three examples suggest that [ŋ] is indeed composed of /ng/, since the properties of nasality and velarity can be split:

(39) spoonerisms involving [ŋ]:

target	observed realization
Bing Crosby	bi**g** cro**n**sby
swing and sway	swi**n** n̩ sway**g**
Springtime for Hitler	spri**g**time for Hi**n**tler

The restricted distribution of [ŋ] in English follows directly from this approach without our having to posit an underlying phoneme /ŋ/ with a defective distribution.

We have modeled the interactions of these two processes as ordered rules, but what about following a constraint-based approach? One of the central claims in most versions of Optimality Theory is that all constraints apply in parallel. The application of all constraints, while potentially highly complex (since *a lot* of constraints are active in any given phonology), should be surface-true, but certain patterns of interaction, such as that seen for English nasals, are **opaque**, that is, the result of a sequence of rules is not transparent or surface-true. Since cases of opacity are predicted not to exist under the simplest set of assumptions within Optimality Theory and yet are well attested across languages, the issue of opacity is of central concern in Optimality Theory (see McCarthy 2007 for a review of these issues and some possible solutions.)

In this section, we have seen a number of ways in which segments might affect each other and evidence for reference to distinctive features, as well as their grouping. We have also seen that the division we made between structure above the level of the segment and subsegmental structure is somewhat artificial, since syllable structure can affect feature specification.

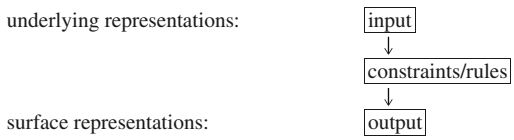
5 Phonology in a Broader Context

In this final section, I summarize the view of phonology sketched here (Section 5.1) and then put the discussion in a broader context (Section 5.2), briefly addressing alternative views of phonology and highlighting some of the current questions that are likely to propel the field forward.

5.1 Phonology as a system

In concluding this introduction to phonology, it is a useful to step back and consider how all the aspects we have discussed fit together. Most basically, a phonology consists of a set of representations – an inventory of sounds, and lexical entries defined by distinctive features matrices – and a system of rules or constraints that act on the representations (Anderson 1985). Fundamental to the generative phonology approach is the idea that the idiosyncratic and predictable information are treated separately: the former is part of underlying representations and the predictable patterns arise through the systematic manipulation of these sounds through rules or constraints. Consider the following schematic figure:

(40)



The underlying representation includes the phonemes for each morpheme in the language and the surface representation incorporates the phonetic variations or allophones, seen in the systematic alternations of the language, introduced as a result of the applications of a system of rules or constraints. The phonological representation includes not only the sequence of sounds, made up of timing units and featural content, but also the hierarchical grouping of these sounds into syllables and higher-level prosodic units. The phonology of a language consists of the whole system taken together. Only by studying the whole phonology of a language, comprising dozens and dozens of rules (or constraints), can we understand its full complexity.

5.2 Emerging trends and research questions

The view presented so far is the general view of generative phonology. While there are interesting and important differences between rule-based and constraint-based views, both of these views

take a similar approach by understanding phonology to be systematically governed behavior which results from the phonological grammar, that is, what a speaker/hearer knows about his/her language – the linguistic competence. There are aspects of phonology that this approach captures elegantly and other aspects that are less well accounted for. In particular there is a tension about the relationship between competence and performance. The dominant view here is that the primary object of interest is competence, which can only be studied through performance (in this case, a language's surface forms).

An alternative view is that competence and performance are inextricably intertwined. In phonological terms, we can consider this in terms of the role of variation. Speech is characterized by variation of all sorts. Variation in pronunciation may be due to differences across styles or registers used by individual speakers or across speakers, dialects, rate of speech, as well as phonological context (the main "stuff" of phonology). A traditional generative approach privileges two kinds of variation: the idiosyncratic differences that capture meaningful distinctions defining sound-meaning relationships of the lexicon, and the systematic or predictable variations defined by phonological contexts (adjacent segments, word position, syllable position, relative prominence, and so forth). Work focusing on other aspects of this variation, especially **variation in the lexicon** and **sociolinguistic variation**, pushes us to see this way of looking at variation as an oversimplification (albeit, I would argue, a very useful oversimplification, see Cohn 2011 in this regard). In this volume see Chapter 12 and Chapter 25.

In this section, we pursue these issues a bit further by considering the following question: What is the relationship between a speaker/hearer's knowledge of their lexicon and knowledge of phonological patterns? Is it the case that the phonology is purely generalizations across the lexicon, or do we "know" the rules or constraint interactions of our language independently?

Pierrehumbert (1994) asks how we can account for the distribution of medial clusters, that is, the fact that certain English consonant sequences are well formed but others are not, e.g., /mpr/, /ndr/ but not */rpm/ or */rdn/. A generative phonology approach, such as that discussed above in Section 3.1, predicts that medial clusters consist of possible codas + possible onsets. On the other hand, if such phonotactic patterns are purely statistical generalizations across the lexicon: "the likelihood of medial clusters is derived from the independent likelihoods of the component codas and onsets" (174)

In a systematic dictionary analysis, Pierrehumbert found roughly 50 monomorphemic medial clusters. The same dictionary listed 147 possible codas and 129 possible onsets. If these could combine freely, we would find 18,963 medial clusters. With some expected restrictions, Pierrehumbert concludes that we could still expect approximately 8,708 medial clusters. She observes "It turned out that almost all the occurring triconsonantal clusters were among the 200 most likely combinations, and that a stochastic interpretation of syllable grammar effectively ruled out a huge number of possible clusters, eliminating the need for many idiosyncratic constraints in the grammar." (169). Pierrehumbert then discusses systematic restrictions that play a role in determining the particular 50 or so medial combinations that are attested among the 200 most likely. She concludes that a probability-based syllable grammar understood in the context of certain more traditional sorts of phonological constraints accounts for the observed patterns.

These results highlight the complex interactions between competence and performance, suggesting that statistical generalizations across the lexicon may be part of how we learn phonology, but once learned, they do indeed constitute patterns of knowledge that are independent from these generalizations. Pierrehumbert (2003) argues that not all phonotactic knowledge is tied to frequency; that is, there is phonological knowledge independent of statistical generalizations across the lexicon. "In light of such results, I will assume, following mainstream thought in linguistics, that an abstract phonological level is to be distinguished from the lexicon proper." (191).

Such integrated approaches to modeling phonological patterns and knowledge fit into a view of **phonology as cognitive science** and have led to the inclusion of new methods, redefining **phonology as an experimental field** – what is termed **Laboratory Phonology** (see Cohn, Fougeron,

and Huffman 2012 for a comprehensive introduction). This includes more extensive work on the multiple facets of variation in sound patterning, as mentioned above. Increasing attention has also been paid to the issue of **the acquisition of phonology**, both in first and second language learning (see Chapter 19 and Chapter 27).

These new approaches and broader perspectives together help advance an understanding of the fundamental question of what speaker/hearers know when they "know" the phonology of their language.

SUGGESTIONS FOR FURTHER READING

Historical Perspective:

Anderson, S. (1985). *Phonology in the Twentieth Century*. Chicago, IL: University of Chicago Press.

Textbooks:

Hayes, B. (2009). *Introductory Phonology*. Malden, MA: Wiley-Blackwell.

McCarthy, J. (2008). *Doing Optimality Theory: Applying Theory to Data*. Oxford: Blackwell Publishing.

Odden, D. (2005). *Introducing Phonology*. Cambridge: Cambridge University Press.

Zsiga, E. (2012). *The Sounds of Language: An Introduction to Phonetics and Phonology*. Malden, MA: John Wiley and Sons.

Handbooks:

Cohn, A., Fougeron, C., and Huffman, M. (2012). *Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press.

de Lacy, P. (2007). *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press.

Goldsmith, J., Riggles, J., and Yu, A. (2011). *The Handbook of Phonological Theory*, 2nd edn. Malden, MA: Wiley-Blackwell.

van Oostendorp, M., Ewen, C., Hume, E., and Rice, K. (2011). *The Blackwell Companion to Phonology*. Oxford: Wiley-Blackwell.

RELEVANT JOURNALS AND SOCIETIES

Phonology (https://www.cambridge.org/core/journals/phonology)

Laboratory Phonology (http://www.labphon.org/home/journal)

The Association for Laboratory Phonology (http://www.labphon.org/)

Manchester Phonology Meeting (http://www.lel.ed.ac.uk/mfm/mfm.html)

NOTES

1 We also talk about the organization of gestures of sign languages as phonology, but we won't develop the parallels here. For further discussion of sign language, see Chapter 18.

2 The description of possible human speech sounds falls within the purview of (linguistic) phonetics. For an introduction, see Ladefoged and Johnson 2011, and Ladefoged and Disner 2012, also Chapter 10.

3 Four consonants following the vowel is not included in (10); an example is *texts* [tɛksts]. In English, most monosyllabic forms with more than two consonants in the coda are morphologically complex, usually involving the [s] or [z] of the plural or third person singular or the [t] or [d] of the past tense.

4 To incorporate this notion of "segmenthood," some approaches include timing units, and others propose an internal hierarchical grouping of features within the segment, including a "root node," as discussed below. An alternative approach to timing is "moraic theory" where the basic units – morae – characterize the weight properties of segments; see Broselow (1995) for a comparison of these approaches.

5 For some speakers, the author included, this is pronounced [sɪŋgɚ], rhyming with *finger* [fɪŋgɚ].

6 For *bomb*, we might assume an underlying /n/ or /m/ or even a nasal consonant that is unspecified for place of articulation.

REFERENCES

Anderson, S. (1985). *Phonology in the Twentieth Century*. Chicago, IL: University of Chicago Press.

Broselow, E. (1995). Skeletal positions and moras. In J. Goldsmith (ed.), *The Handbook of Phonological Theory* (pp. 175–205). Oxford: Blackwell.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English* (*SPE)*. New York: Harper and Row.

Clements, G. N. (2003). Feature economy in sound systems. *Phonology*, 20(3), 287–333.

Clements, G. N. and Hume, E. (1995). The internal organization of speech sounds. In J. Goldsmith (ed.), *The Handbook of Phonological Theory* (pp. 245–306). Cambridge: Blackwell.

Cohn, A. (2011). Features, segments, and the sources of phonological primitives. In G. N. Clements and R. Ridouane (eds), *Where Do Features Come From?* (pp. 15–41). Amsterdam: John Benjamins.

Cohn, A., Fougeron, C., and Huffman, M. (2012). *Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press.

Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27–52.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago, IL: University of Chicago Press.

Hyman, L. (2008). Universals in phonology. *The Linguistic Review*, 25, 83–137.

Kager, R. (1999). *Optimality Theory*. Cambridge: Cambridge University Press.

Kenstowicz, M. and Kisseberth, C. (1979). *Generative Phonology: Description and Theory*. New York: Academic Press.

Ladefoged, P. and Disner, S. (2012). *Vowels and Consonants*, 3rd edn. Oxford: Wiley-Blackwell.

Ladefoged, P. and Johnson, K. (2011). *A Course in Phonetics*, 7th edn. Boston, MA: Wadsworth.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maddieson, I. and Precoda, K. (1990). Updating UPSID. *UCLA Working Papers in Phonetics*, 74, 104–111.

McCarthy, J. (2007). Derivations and levels of representation. In P. de Lacy (ed.), *The Cambridge Handbook of Phonology* (pp. 99–117). Cambridge: Cambridge University Press.

McCarthy, J. (2008). *Doing Optimality Theory: Applying Theory to Data*. Oxford: Blackwell Publishing.

Mielke, J. (2008). *The Emergence of Distinctive Features*. Oxford: Oxford University Press.

Pierrehumbert, J. (1994). Syllable structure and word structure: a study of triconsonantal clusters in English. In P. Keating (ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III* (pp. 168–88). Cambridge: Cambridge University Press.

Pierrehumbert, J. (2003). Probabilistic phonology: discrimination and robustness. In R. Bod, J. Hay, and S. Jannedy (eds), *Probabilistic Linguistics* (pp. 177–228). Cambridge, MA: The MIT Press.

Zec, D. (2007). The syllable. In P. de Lacy (ed.), *The Cambridge Handbook of Phonology* (pp. 161–93). Cambridge: Cambridge University Press.

11 Morphology[1]

ANDREW SPENCER

1 Introduction

Morphology is about the structure of words. In a language like English many words have an internal structure, consisting of one or more **morphemes**. Thus, the form *cats* comprises the root morpheme 'cat' to which is added the suffix morpheme 's' indicating plural. Why do we say that the expression *the cat* or its French equivalent *le chat* is two separate words, rather than take *the/le* to be prefixes, especially given the French expression *l'ami* 'the friend'? If the 'l' of *l'ami* is a word, why can't we say that the 's' of *cats* is also a word? Here we need the help of syntax: *the cat* is a phrase which can be extended by the addition of other phrases: *the very black cat* and French *l'ami* behaves similarly. The form *cats* can never be split up this way, the reason being that the 's' component is an element which can only exist as part of a word, specifically at the end of a noun. In other words, 's' is a suffix and hence a **bound morpheme**. The property of indivisibility exhibited by *cats* is **lexical integrity**. A single word such as *cats* contrasts rather neatly with the fully fledged (but synonymous) phrase *more than one cat*, in which it is clear that *more*, *than*, and *one* are all independent words and can all be separated by other words or phrases.

This chapter will examine the different structures that words exhibit and the morphological relationships they bear to each other, and the nature of the morpheme. We begin by clarifying the notion "word" itself.

1.1 The lexeme concept

If we ask how many words are listed in (1) we can give at least two answers.

(1) {cat, cats}

In one sense there are obviously two, but in another sense there is only one word, CAT, and only one entry will be found in a dictionary for it. The plural, *cats*, is formed by a completely general rule from the singular form *cat* and there is no need to record the plural form separately. In addition, we can describe *cat* as "the singular form of the word CAT" and *cats* as "the plural form of the word CAT." On the other hand, the singular form of the word SHEEP has exactly the same form

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

as the plural, namely *sheep*, even though, in another sense, these are two distinct words, namely "the singular of SHEEP" and "the plural of SHEEP."

It is rather useful to have different terms for these three different senses of the word "word." We will therefore say that there is a **lexeme** CAT which has two **word forms**, *cat* and *cats*. The names of lexemes are conventionally written in small capitals. The grammatical description "the singular/plural of CAT" is a **grammatical word**. Thus, *sheep* is one word form corresponding to one lexeme, SHEEP, but it is two grammatical words (singular/plural of SHEEP).

We can think of a lexeme as a complex representation linking a (single) meaning with a set of word forms (or, to be very strict, linking a meaning with a set of grammatical words, which are then associated with corresponding word forms). From the point of view of the dictionary (or lexicon), this representation is therefore a **lexical entry**. If several sets of forms correspond to one meaning we have pure **synonymy**: e.g., {*boat*, *boats*}, {*ship*, *ships*}. If a single form corresponds to more than one completely unrelated meaning, as with {*write*, *right*, *rite*}, or {*bank*, *bank*}, then we have **homophony** or **homonymy**. Homophones/homonyms are distinct lexemes which happen to share the same shape (written and/or spoken). In some cases these meanings are felt to be related to each other, and dictionaries tend to treat this as an instance of **polysemy**. Thus, the word "head" means a body part, the person in charge of an organization, a technical term in linguistics, and so on, and these meanings are associated by some kind of metaphorical extension. However, it is better to think of this type of relatedness as homophony (but see Section 6.2 for an example of systematic polysemy with verbs such as BREAK).

A pairing of form with meaning is a **sign**, of which the lexeme is a prototypical example. The traditional definition of morpheme is "the smallest meaningful component of a word," and this entails that we consider all morphemes as signs. However, this turns out to be very controversial, for some types of morpheme, at least.

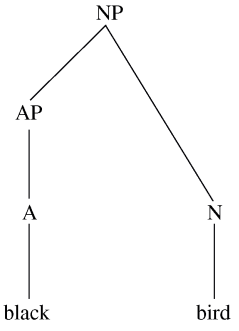
1.2 Types of word formation: inflection, derivation, compounding

It is common to distinguish **inflection**, in which we create word forms of lexemes, such as the plural or past tense, **derivation**, in which we create new lexemes from old lexemes, and **compounding**, in which a single word is formed by combining two other words. We begin with compounds.

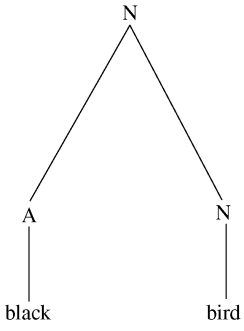
The most straightforward type of compound consists of two concatenated words: *morphology* + *article* = *morphology article*; *house* + *boat* = *houseboat*. The right-hand member is the **head** of the compound, determining the syntactic category and meaning of the whole (a morphology article is a kind of article, a houseboat is a kind of boat, as opposed to a boathouse, which is a kind of house). The left-hand member is the modifier. In transparent cases such as *morphology article* the meaning of the whole is derived from the meanings of the components, though the precise meaning is indeterminate and depends on the context of use.

There is an important distinction in many languages between compounds and phrases. In many cases the difference is obvious. In a hackneyed example such as the compound *blackbird* as opposed to the phrase *black bird*, the compound has stress on *black*, while the phrase is stressed on *bird* (in neutral contexts at least). Moreover, a black bird is necessarily black, while a blackbird is a particular species of bird whatever its color. This means that the semantics of *blackbird* is **noncompositional**, i.e., we can't determine the meaning of the whole just from the meanings of the parts. The semantics of phrases (idioms apart) is compositional. The difference can be illustrated syntactically as in (2, 3) (making very conservative assumptions about syntactic structure):

(2)

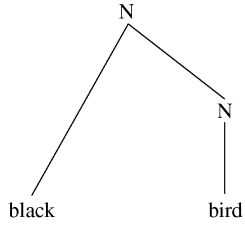


(3)



This is the standard story, though there are interesting subtleties. First, there is no way of determining the syntactic category of the modifier in blackbird, because it is fixed as part of the compound and can't be subjected to any of the morphological or syntactic manipulations of real adjectives. Thus, we can't say *an apparently blackbird, though we can say an apparently black bird. Moreover, since black doesn't actually mean "black" in blackbird (female blackbirds are brown), the modifier black has no meaning either, it is just a bare morphophonological shape. Therefore, (3) should be rewritten as (4):

(4)



The point is that *blackbird* is a lexicalized compound whose internal structure is only of historical significance, unlike a nonlexicalized coinage such as *morphology article*. In time, with changes in pronunciation, the compound structure may become opaque. Thus, *husband* is derived etymologically from (modern) "house" and "bond," but it isn't recognized as a compound by anyone except students of Middle English.

Noun+noun compounding is a fully **productive** process in English. This means that it applies freely in principle to all the lexemes of the language of the relevant type, allowing new forms to be created at will even if they have never been used before. Such processes therefore have to be semantically regular, without any lexicalized idiosyncrasy of meaning, otherwise, hearers would have no way of knowing what a new coining was supposed to mean (see Bauer 2001). The meaning of such compounds is admittedly vague: a morphology article is an article which has some connection with morphology. On the other hand, adjective+noun compounds aren't productive and there are very few compounds in English like *swearword* or *drawbridge* consisting of verb+noun.

There is a great variety of compound types in the languages of the world (Bauer 2009). One widespread type lacking in English is known as noun incorporation (see Mithun 1984). In the most common type the noun expressing the direct object of a verb may form a compound with the verb's root. In (5) we see two examples from Chukchi (Northeast Siberia):

(5)	a.	Gəmnan	tə-piri -gʔen	pojgə-n
		I.ERG	1SG.SUBJ-take-3SGOBJ	spear-ABS
	b.	Gəm	tə-pojgə-pere-gʔak	
		I.ABS	1sgSUBJ-spear-take-1sgSUBJ	
		'I took the spear'		

In (5a) the subject pronoun is in the ergative case (the case used to mark the subject of a transitive sentence), while the object is in the absolutive case. Being transitive, the verb agrees with both the subject and the object. In (5b) the root of the object noun has formed a compound with the verb root. This renders the verb intransitive, so it agrees solely with the first-person subject. The subject pronoun is now in the absolutive case, the case used for intransitive subjects. Finally, notice that the 1SG prefix comes to the left of the incorporated noun root and the vowels of the root have changed, due to vowel harmony, under which the "weak" vowel /i/ is changed to /e/ when there is a "strong" vowel elsewhere in the word (e.g., the /o/ of *pojgə*). Vowel harmony only operates within a word, and this helps us identify the incorporative complex as a single word form morphologically. Examples (5a,b) differ slightly in emphasis but are otherwise synonymous. Thus, it is clear that *pojgə* still realizes the 'spear' lexeme even when it is compounded. Noun incorporation is completely productive in Chukchi, with very few restrictions.

Turning to derivation, the nouns *writer, painter, walker* are clearly related to the verbs *write, paint, walk*, meaning roughly "person who writes, paints, walks," by suffixation of *-er*. I shall call these **subject nominals**. It is customary to treat *write* and *writer* as distinct lexemes related by derivation, rather than word forms of a single lexeme. For instance, *writer* is a noun, while *write* is a verb. The morphological operations which realize derivation (such as *-er* affixation) may or may not be regular and productive. Thus, APPLY has a subject nominal *applic-ant*, with irregular suffix *-ant* added to an irregular form of the root, *applic-*. I discuss derivation in more detail in Section 6.1.

As a verb lexeme, WRITE has its own set of grammatical words expressed by the forms *write, writes, writing, wrote, written*. Similarly, WRITER has its own set of forms: *writer, writers*. These grammatical words are the **inflected** forms of the lexeme and the process of constructing inflected forms is known as inflection ("inflectional morphology"). The meanings of the inflected form are predictable (plural of noun, past tense of verb, or whatever), while the shape of inflected forms is generally determined by affixation to the **stem** form of the lexeme. The stem consists of the root and any derivational affixes. In morphologically complex languages a given lexeme might have several stems for different types of inflection (e.g., all verbs may have separate present tense and past tense stems). Irregularity, either in the stem or the affix, is not uncommon. Thus, *knife* has the irregular stem form *knive-* in the plural (*knives*), while *ox* has the irregular suffix *-en* (*oxen*). Irregularity of form can be complete as in **total suppletion**, when one inflected form bears no shape relation to the rest of the paradigm (e.g., *went* as the past tense of GO). Where there is still some overlap we talk of **partial suppletion** (as in *brought ~ bring*, where the first two consonants are identical). Even where the shapes are irregular, the past tense meaning is exactly the same as it is for any other verb, whether irregular (such as *write ~ wrote, bring ~ brought, go ~ went*) or regular (e.g., *scribble ~ scribbled*).

Inflections express **grammatical** or **functional categories**. The inflectional system organizes the forms of words into systematic groupings or **paradigms**. There are essentially two sorts of function subserved by inflection. Many inflections signal an aspect of meaning which is **grammaticalized**, such as number (singular vs. plural) or tense. This means that the words of a given class obligatorily signal the grammatical distinction. Thus, all verbs in English have to have a past tense (even if these are not actually distinct forms, as in *put*). Booij (2007) calls this type of morphology **inherent inflection**.

One typical inherent inflection for nouns is **case**, in which the grammatical or semantic role of a noun in a sentence is shown by its form. As we can see in (6), a Russian noun generally has distinct forms for the subject (*Lena*), direct object (*knigu*), or indirect object (*Ire*). These are case-inflected forms respectively of the lexemes LENA, KNIGA, IRA.

(6) Len-a dala Ir-e knig -u
Lena-NOM gave Ira-DAT book-ACC
'Lena gave Ira a book'

Verbs generally exhibit much greater variety in their inflectional systems. Two common inherent inflections are **tense** and **aspect**. Tense refers to anchoring in time, as with English *wrote* (past) as opposed to *writes* (non-past – present or future reference). A given language may distinguish a number of different tenses (such as recent vs. remote past) or no tense at all. Aspect refers to the manner in which an event unfolds over time. A very common aspectual distinction is that between completed (perfective) and noncompleted (imperfective) events. In Slavonic languages most verbs have separate perfective and imperfective paradigms, e.g., *op'isat'* (perf.) ~ *op'isivat'* (impf.) 'describe' (see also Section 6.2). Many languages have very rich aspectual markings modifying the meaning of the base verb in very subtle ways. Below is just a small selection of the 15 aspectual affixes described for Chukchi by Skorik (1977: 179–202):

(7) -lʔet- prolonged continous action:
ʔəttʔe ninepiŋku-**lʔet**-qin ottəlgən
dog jump-**ASP**-3/3 stick
'The dog jumped over the stick **over and over again**'

(8) -cir- prolonged interrupted action:
ŋinqejmuri nʔejŋew-**cir**-muri jaralʔa
us.children called-**ASP**-1plOBJ people.at.home
'The people at home **kept** calling (to) us children'

(9) -cit/cet- alternating action
natcə-**cet**-qenat
hide-**ASP**-3plSUBJ
'They played at hide-and-seek'

(10) -sk əcet- accelerated action
qənwer ŋəto-**sqəcat**-gʔe gəmnin tələtumgin
at last come.out-**ASP**-3sgSUBJ my companion
'At last my companion sprang out'

More than one of these can be combined:

(11) mət-ra-təla-tenmawə-plətko-ŋŋo-gʔa
1pl-FUT-**GRADUALLY**-prepare-**FINISH-BEGIN**-FUT
'we will **begin to gradually finish** the preparations'

Other types of verb inflection include mood (whether a statement is presented as fact, possibility, hypothetical situation, and so on), including the subjunctive mood of Romance languages, the optative expressing a wish (e.g., Ancient Greek), the imperative for issuing commands, and

the interrogative, a special set of verb forms used for asking questions (as found in the Eskimo languages). Many language groups signal polarity (i.e., negation) inflectionally (Bantu, Turkic, Athapaskan, and others). It is very common for a single inflectional affix to realize a complex combination of tense, aspect, mood, and polarity information.

Any of the above functional categories can be expressed syntactically, by word order or by function words such as the English aspectual auxiliaries (*has been reading*). One purely morphological type of inherent inflection is **inflectional class**: **declensions** for nouns and adjectives and **conjugations** for verbs. Which noun or verb goes in which class is in general arbitrary. Russian nouns (Table 11.1) can be put into four main declensions depending on the inflections they take.

I have given two subtypes of Class I nouns, one animate, the other inanimate. In the inanimates the accusative case is always the same as the nominative, while in the animates the accusative takes the form of the genitive. This type of situation, in which parts of a paradigm are systematically identical, is known as **syncretism**. There are other syncretisms here, too. For instance, the dative, instrumental, and prepositional plural endings are the same for all classes, that is, the class feature is neutralized and there is effectively a single set of endings for the whole of the class 'noun'. On the other hand, the behavior of pairs of lexemes such as LAW and BOY require us to set up a covert category of animacy for Russian, which never has any direct expression (there is no form which has a suffix identifiable as the "animacy" suffix) but which is nonetheless part of the inflectional system. Note that it is the property "animacy" which is covert, not the accusative case. We know this because Class II nouns have a separate accusative, in the singular at least (Corbett and Fraser 1993).

Russian verbs (Table 11.2) inflect so as to indicate the person/number of their subject (see below on "agreement") as well as for tense, and occur in two main conjugations.

As can be seen, the endmost suffixes are common to both classes, except in 1sg, 3pl forms. Both types have a special stem-forming suffix, *-aj-* and *-i-* respectively, and Class I has in addition a "linking vowel" *-o-*. The *-aj/-i* formatives are found throughout the inflectional system of the verbs.

Table 11.1 Russian noun classes.

	Class I inanimate	Class I animate	Class II	Class III	Class IV
	'law'	'boy'	'room'	'bone'	'place'
Singular					
Nominative	zakon	mal´čik	komnata	kost´	mesto
Accusative	zakon	mal´čika	komnatu	kost´	mesto
Genitive	zakona	mal´čika	komnati	kost´i	mesta
Dative	zakonu	mal´čiku	komnate	kost´i	mestu
Instrumental	zakonom	mal´čikom	komnatoj	kost´ju	mestom
Prepositional	zakone	mal´čike	komnate	kost´i	meste
Plural					
Nominative	zakoni	mal´čiki	komnati	kost´i	mesta
Accusative	zakoni	mal´čikov	komnati	kost´i	mesta
Genitive	zakonov	mal´čikov	komnat	kostej	mest
Dative	zakonam	mal´čikam	komnatam	kost´am	mestam
Instrumental	zakonam´i	mal´čikam´i	komnatam´i	kost´am´i	mestam´i
Prepositional	zakonax	mal´čikax	komnatax	kost´ax	mestax

(The symbol ´ represents palatalization. Consonants are always palatalized before /e/. The case names are traditional and represent a variety of syntactic functions.)

Table 11.2 Principal Russian verb classes.

	Non-past tense	
	Class I	Class II
person/number		
1sg	uzn-aj-u	govor´-u
2sg	uzn-aj-o-š	govor´-i-š
3sg	uzn-aj-o-t	govor´-i-t
1pl	uzn-aj-o-m	govor´-i-m
2pl	uzn-aj-o-te	govor´-i-te
3pl	uzn-aj-ut	govor´-at

Class I verb *uznat´* 'to recognize', Class II verb *govor´it´* 'to speak'.

The other role of inflection is to realize the syntactic functions of **agreement** and **government**. This is what Booij (2012) calls **contextual inflection**, because it is determined by the syntactic context in which the lexeme is used. In many languages a verb must **agree with** its subject and/or object, by crossreferencing various of their properties. This occurs marginally in English for third-person non-past verb forms: *Harriet writes* vs. *the girls write*. In Chukchi, transitive verbs agree with both the subject and the object, in rather complex ways. The system for one of the six tense forms in the indicative mood is shown in Table 11.3 (cf. Muravyova 1998; empty cells represent nonexistent forms in which the subject and object would have the same-person features):

Table 11.3 Chukchi transitive verb *pela-* 'to leave (someone, something)', simple past tense.

Subj	Obj		Subj	Obj	
1sg	1sg	_____	1pl	1sg	_____
	2sg	tə-pela-gət		2sg	mət-pela-gət
	3sg	tə-pela-gʔan		3sg	mət-pela-gʔan
	1pl	_____		1pl	_____
	2pl	tə-pela-tək		2pl	mət-pela-tək
	3pl	tə-pela-nat		3pl	mət-pela-nat
2sg	1sg	ena-pela-gʔe	2pl	1sg	ena-pela-tək
	2sg	_____		2sg	_____
	3sg	pela-gʔan		3sg	pela-tkə
	1pl	pela-tko-gʔe		1pl	pela-tko-tək
	2pl	_____		2pl	_____
	3pl	pela-nat		3pl	pela-tkə
3sg	1sg	ena-pela-gʔe	3pl	1sg	na-pela-gəm
	2sg	na-pela-gət		2sg	na-pela-gət
	3sg	pela-nen		3sg	na-pela-gʔan
	1pl	na-pela-mək		1pl	na-pela-mək
	2pl	na-pela-tək		2pl	na-pela-tək
	3pl	pela-nenat		3pl	na-pela-nat

The verb references the person and number both of the subject and of the object, though there is no simple relationship between many of the affixes and their functions. Thus, although the prefixes *tə-* and *mət-* clearly mean '1sg/1pl subject' respectively, the prefix *na-* seems to mean '3pl subject' or '3sg subject with 2nd person object or 1pl object' and the suffix *-nen* seems to mean '3sg object but only if the subject is 3sg'. One consequence of this is that some forms correspond to more than one subject-object pairing. For example, *napelagət* means either '3sg leaves 2sg (s/he leaves thee)' or '3pl leaves 2sg (they leave thee)'. The system proves to be even more complex than this when the full set of tenses, moods, and voices is taken into account. Patterns such as Chukchi verb inflection are typical of languages with rich agreement systems, and such data have been instrumental in changing the views of linguists about the nature of the morpheme.

Adjectives often agree with the nouns they modify. In Russian, an adjective agrees with its noun in number and case:

(12)	a.	bol'šoj mal'čik 'big boy'	masculine nominative singular
	b.	bol'šogo mal'čika	masculine genitive singular
	c.	bol'šim mal'čikam	masculine dative plural
(13)	a.	bol'šaja devuška 'big girl'	feminine nominative singular
	b.	bol'šoj devuški	feminine genitive singular
	c.	bol'šim devuškam	feminine dative plural

It might be thought that the adjective agrees in declension, but this is wrong. All nouns in Russian have one of three **genders**, masculine, feminine, or neuter. Male and female humans are masculine and feminine respectively and for other nouns gender depends largely on declensional class. Members of Class I are masculine, those of Classes II and III are feminine, and those of Class IV are neuter. However, there are certain exceptions. Thus, the word *mužčina* 'man' belongs to Class II, yet it is masculine: *bol'šoj mužčina* 'big man'. As is stressed by Aronoff (1994), gender is an essentially syntactic property, which governs agreement. Declension class is a purely morphological property to which the syntax has no direct access. Aronoff points out that the existence of arbitrary inflectional classes is one of the prime motivations for treating morphology as an autonomous linguistic module.

We have seen that a direct object in Russian is in the accusative case. This can be thought of as an instance of **government**: a transitive verb **governs** the accusative. Likewise, prepositions in Russian have to take specific cases, as shown in (14):

(14)	a.	okolo near 'near the house'	dom-a house-GEN
	b.	v in 'into the house'	dom house.ACC
	c.	v in 'in the house'	dom-e house-PREP

Notice how "motion toward" as opposed to "location at" is signaled solely by case choice in (14b,c), otherwise, it is an arbitrary matter which preposition governs which case.

One of the perennial theoretical questions in morphology is whether there is a clearcut distinction between inflection and derivation and if so how to draw it. Inflection is often thought

to be "of relevance to syntax," which is clearly true of contextual inflection, but not so obvious with inherent inflection. Yet we don't want to say that plurals or past tenses are derivational and hence create new lexemes. Booij's contextual/inherent distinction is designed to ameliorate this problem (though we are now left with the task of distinguishing inherent inflection from derivation). A typical borderline case is that of the aspectual forms of Chukchi given above. Chukchi has a set of six tense–aspect forms in which aspect (roughly perfective vs. imperfective) is grammaticalized and expressed as part of the obligatory conjugation system. However, the affixes illustrated in (7–11) are not like this. Rather, they are optional elements which are added to modify the overall meaning of the verb. Does this make them derivational, then? Do we wish to say that 'to *Verb* in a prolonged interrupted fashion' is a new lexeme related to *Verb* (derivation) or an inflected form of the lexeme VERB (inherent inflection)? Cases like this are quite common and promise to provide fertile ground for future research into the problem.

1.3 Other types of realization: clitics and periphrases

Although many morphologists regard morphology as in principle autonomous of other aspects of grammar, this doesn't mean that it is always possible to draw a clear dividing line. In particular, morphology shades into syntax, in the sense that there are processes which share some of the properties of each. We have already seen compounds in English. The lexicalized compounds such as *blackbird* are objects of morphology in that they behave just like individual words, not like syntactic phrases, but other compound types, such as Chukchi noun incorporation, have syntactic properties, such as realizing the direct object of a verb. Within inflection there are two respects in which morphological and syntactic organization systematically interact: the phenomena of **clitics** and of **periphrasis**.

A clitic is a word which has lost some of its syntactic independence and which behaves in some ways like an affix. Usually, a clitic is prosodically weak, for instance, it may lack stress or tone, so that it has to be phonologically attached to a **host**. When it attaches to the left of the host, comparable to a prefix, it is a **proclitic**, and when it attaches to the right, like a suffix, it is an **enclitic**. The Daghestan language Udi is said to have **endoclitics**, which come inside the host. English has clitic forms of the auxiliary verbs BE and HAVE, some of which are so affix-like that they consist just of one consonant: *'s* in *Tom's a linguist, Harriet's left* (cf. *Tom is a linguist, Harriet has left*). Clitics can come together to form a string often known as a **clitic cluster**: *Dick'll've left* (cf. *Dick will have left*). In most languages the order of clitics is strictly fixed even if word order is normally very free in that language.

Clitics express all the properties that are commonly expressed by inflections, such as case, number, definiteness (for nouns), tense, mood, aspect (for verbs), as well as expressing pronominal arguments (possessors for nouns and subject or object arguments for verbs, for instance). In addition, discourse particles corresponding to English *well, after all, y'know,* and many others are often expressed as clitics, as are formatives realizing questions or negation.

In some languages the clitic or clitic cluster occupies a syntactic position reserved just for clitics. A very common variant is the Wackernagel position (named after the Indo-Europeanist who first described the phenomenon in detail for Greek and Latin in 1892). This is the position immediately after the first word, or the first phrase, of a clause. In Czech, a clitic cluster can include an auxiliary verb such as *jste* '(you) are', pronouns (in a fixed order) such as *si ho* 'to-oneself him/it', and finally a discourse clitic such as *prý* 'apparently, so they say': *Vy jste si ho prý vzeli* 'you took it for yourselves, they say'. Bosnian-Croatian-Serbian (sometimes still called "Serbo-Croat") has a Wackernagel clitic system very similar to that of Czech (they are both Slavic languages), but unlike Czech it optionally allows the clitic cluster to appear inside the first phrase of the clause, taking as its host the first (prosodically complete) word of that phrase. For instance, in (15b) the cluster breaks up the noun phrase 'that poet' by attaching to the demonstrative *taj* 'that' as its host:

(15) a. Taj pesnik *mi* *je* napisao knjigu
that poet *for-me* *AUX* wrote book

b. Taj *mi* *je* pesnik napisao knjigu
that *for-me* *AUX* poet wrote book
'That poet wrote me a book'

Elements which are often referred to as clitics in standard grammars sometimes have more of the properties of affixes than of the clitics we have seen in English, Czech, or Bosnian-Croatian-Serbian. Romance languages such as French, Italian, and Spanish typically have weak pronoun forms which cluster around the verb. These are usually referred to as clitics, but in most cases they can equally be thought of as loosely attached affixes. In European Portuguese the pronominals behave like clitics when placed to the left of the verb but have all the properties of affixes when placed to the right of the verb. A recent survey of clitics and recent research on clitics can be found in Spencer and Luís (2012).

If a language doesn't express functional notions by means of inflectional morphology it will typically do so by means of function words (some of which may be clitics, of course). Such function words often have syntactic idiosyncrasies of their own, as is the case with the English auxiliary verbs. Nonetheless, it is usually assumed that a string of auxiliary verbs such as *would have been* in *Harriet would have been reading* is governed by syntactic and not morphological principles. However, we sometimes find cases in which a piece of syntax seems to fulfil a role which we would otherwise expect to be fulfilled by inflected morphological forms. To distinguish such cases from ordinary syntax we often speak of "periphrastic constructions," or "periphrases." (We also often speak of **analytic** expressions as opposed to purely morphological or **synthetic** expressions.)

A case in point, which illustrates the theoretical significance of periphrasis, is the Latin perfective passive. Latin (transitive) verbs inflect for, inter alia, person/number, present~past~future tense, perfective~imperfective aspect, indicative~subjunctive mood, and active~passive voice. A sample of these categories for the verb AMO 'to love', inflected for a 3sg subject, is shown in Table 11.4 (using standard orthography).

Given this partial paradigm we would expect to see the empty cells for the passive of the present and past perfective filled by the forms **amavitur*, **amaveratur*. Phonologically (and, indeed, morphologically) there seems no reason why these forms shouldn't exist. However, what we actually find for the whole of the perfective passive subparadigm is a periphrastic construction, consisting of the perfective passive participle (an adjectival form), and an appropriate form of the verb BE: *amatus est/erat* 'he had/had been loved'. In nearly all uses this combination clearly consists of two distinct words, which can easily be separated (Latin had very free word order).

Interestingly, the normal form of BE that is used in this periphrasis in the classical period is a form which elsewhere has imperfective meaning, despite the fact that the periphrase as a whole

Table 11.4 Partial paradigm for Latin AMO 'love'.

	ACTIVE	PASSIVE
present indicative	amat	amatur
present subjunctive	amet	ametur
future	amabit	amabitur
past imperfective	amabat	amabatur
present perfective	amavit	
past perfective	amaverat	

realizes perfective. There is a perfective past form of BE, *fuit*, but that is not usually found in this periphrase. All the perfect forms that are expressed morphologically (synthetically) in the active form are expressed periphrastically in the passive, e.g., *amatus sit* (present subjunctive), *amatus esset* (imperfect subjunctive).

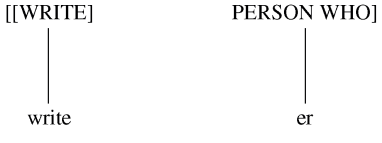
The Latin perfective passive illustrates two characteristic properties of periphrases. First, it is quite typical for a periphrase to be idiomatic or noncompositional in the sense that it is not possible to predict the exact meaning of the periphrase from the meanings of its parts. The clash between imperfective *erat* and the perfective meaning of *amatus erat* illustrates this. Second, and this can be thought of as part of the definition of a periphrasis, the periphrase occupies a cell which we would otherwise expect to be filled by a single inflected (synthetic) word form. In other words, the periphrases are found at the intersection of the passive subparadigm with the perfective subparadigm. In their discussion of the properties of periphrasis Brown et al. (2012) take this property as effectively criterial for being a genuine periphrase.

2 The Morpheme Concept and Agglutinating Morphology

2.1 Item-and-Arrangement morphology

If we return to the example of *writer* we can easily segment it into two component forms or **morphs**, a verb **base**, *write*, and a derivational suffix *-er*. (I use "base" to mean any form of a lexeme to which an affix is added, whether derivational or inflectional). It is usually claimed that the suffix as well as the base has a meaning and that the meaning of the derived word is obtained by combining the meanings of the two component morphs as shown in (16):

(16)



On this basis both of the morphemes are a pairing of a pronunciation (or shape, the morph) and a meaning. They are thus signs and hence are both lexemes, making the combination essentially a compound, like *houseboat*. Admittedly, *-er* is a bound morpheme, but in many languages lexemes can be compounded in the form of a bound stem. Thus, the form *pojgə* in the Chukchi noun incorporation example (5b) is in fact a bound stem form (the word for "spear" itself always surfaces with a case suffix), and even in English one might argue that there are compounds consisting solely of bound roots, the so-called neoclassical compounds such as *gram-o-phone* (or *phon-o-gram*).

The traditional account of plural morphology treats the plural suffix in the same way, a type of sign with a phonology and a semantics, as shown in (17):

(17) *-z* = </z/, [plural]>

This way of looking at things immediately leads us to the conclusion that words have a hierarchical structure which can be represented as a tree diagram. A possible structure for *writers* is shown in Figure 11.1.

In Figure 11.1 the grammatical property [plural] is said to **percolate** from the suffix to the top of the tree, ensuring that the entire word is interpreted as a plural form. This type of approach is often referred to as the "word syntax" model (Toman 1998).

The set of assumptions underlying representations such as Figure 11.1 derive from what is generally called the **Item-and-Arrangement** theory (IA for short): morphemes are "things" which are arranged in a particular way ("morphotactics") and which contribute their meaning to

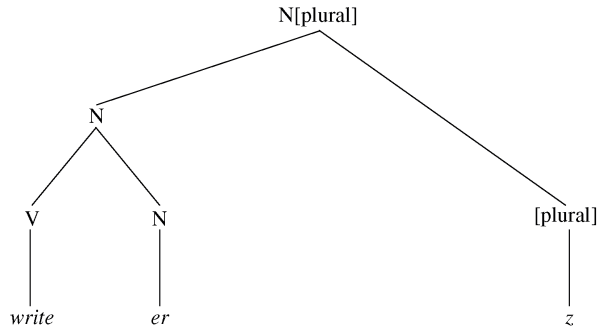


Figure 11.1 Hierarchical analysis of *writers*.

the meaning of the whole word. In an "ideal" morphological system each morpheme contributes one meaning and each meaning is associated with just one morpheme ("one form – one function"). Such a morphological ideal is often called **agglutination** (and morphologists still sometimes speak of "agglutinating" languages where this type of morphology predominates).

It should be obvious that this approach is at odds with the lexeme concept: the plural form *cats* would not, after all, be a word form belonging to an abstract lexeme, CAT, rather it would be a compound form, in which the meaning of the suffix (or perhaps we should say the head of the compound?), PLURAL, is grammaticalized. Where inflection is concerned this has proved impossible to maintain, for three main reasons. First, it is not always possible to identify a single segmentable morph for the putative morpheme, for instance, where is the plural morpheme in *men* (see subsection 2.2.2)? Second, there are significant deviations from the form–meaning pairing in affixation and these undermine the assumption that inflections are signs. Third, for such a theory to work we must be able to explain in a satisfactory way how complex words are constructed, and in particular how the morphemes get strung out in the right order. For complex inflectional systems this turns out to be very tricky.

2.2 *Deviations from agglutination*

The "ideal" type of morphology is often seen as the addition of a semantically transparent affix to a base, so-called **concatenative** morphology. This is what Corbett (2010) calls canonical morphology. There are several ways in which morphological systems present deviations from the agglutinating canonical pattern of one form – one function. The first is found wherever a morpheme can or must appear in more than one phonological shape depending on its surroundings (allomorphy). Beyond this, we find that there are operations which can't easily be analyzed as the addition of a meaningful element but rather take the form of a phonological process, often called **nonconcatenative** morphology. Languages abound in such operations and there have been a number of ingenious ways of dealing with them. I shall mention just three particularly salient cases here.

2.2.1 Allomorphy The regular past tense ending appears as three different morphs in English, depending on the final sound of the noun stem: *walk-ed* (/t/), *jogg-ed* (/d/), *trott-ed* (/əd/, where /ə/ is the schwa or reduced vowel). This variation is **allomorphy**, and we say that /t, d, əd/) are the three **allomorphs** of the past tense morpheme. In this case the allomorphy is conditioned solely by the phonology of the stem: /əd/ after /t, d/; /t/ after a voiceless sound; /d/ elsewhere. However, other cases of allomorphy may be phonologically **irregular**. For instance, while *mend*

and *pen* have regular past tense forms (*mended*, *penned*), the verb *bend* takes an unexpected *-t* ending and adds this to an irregular stem form lacking the final *-d*: *ben-t*. Thus, both stem and suffix show irregular (unpredictable) allomorphy. Where a given morpheme is realized by more than one allomorph we have a (mild) deviation from the agglutinative ideal.

2.2.2 Processual morphology Certain types of irregular verb in English form their past tense by taking the basic root, *sing, run, drive, write*, and changing its vowel: *sang, ran, drove, wrote*. This kind of process is called **ablaut** or **apophony**. In a number of languages, most famously Semitic languages such as Arabic and Hebrew, apophony is regular and widespread throughout the grammar. It is very difficult to represent this in terms of the addition of an affix to a base (though see McCarthy 1982 for the classic Item-and-Arrangement analysis of Semitic).

Another well-attested phenomenon is **reduplication**, illustrated by the Tagalog examples in (18):

(18)	a.	sulat	'writing'	su-sulat	'will write'
	b.	basa	'reading'		
		mambasa	infinitive	mam-ba-basa	nominalization
	c.	magpa-sulat	causative	magpa-pa-sulat	'will cause to write'

Here, morphological categories are signaled by a kind of prefix, which consists of a copy of certain segments of the stem. Any analysis of this phenomenon has to recognize that there is a process involved at some level.

A particularly drastic type of nonaffixal morphology is so-called **subtractive** morphology in which a morphological category is signaled by loss of a portion of the base. Anderson (1992: 64–6) lists a number of inflectional processes which, apparently, have to be so analyzed, such as the example in (19) from the Muskogean group:

(19)	a.	balaa-ka 'lie down (sg.)'	bal-ka 'lie down (pl.)'	(Alabama)
	b.	bonot-li 'roll up (sg. Obj.)	bon-li 'roll up (pl. Obj.)	(Choctaw)
	c.	atakaa-li 'hang (sg.)'	atak-li 'hang (pl.)'	(Koasati)

Here, the plural or plural object form of the verb is derived from the singular form by removing the rhyme of the final syllable of the stem: *bal<aa>*, *bon<ot>*, *atak<aa>*.

2.2.3 Form–meaning deviations In this subsection we examine the idealization that one form corresponds to one meaning/function and vice versa. We already know of two types of deviation: synonymy (many forms – one meaning) and homonymy (one form – many meanings). However, four additional types of deviation can be distinguished when we look at the meanings or functions of morphemes within a single word.

1. *One morph, two meanings*

   The Russian case system shown in Table 11.1 clearly has a grammatical category of 'plural' but no single identifiable morpheme signaling number. Thus, *-am* means 'dative' and 'plural' simultaneously. Note that this is not homonymy, because the suffix simultaneously conveys both meanings within the same word form and these meanings are inseparable. We say that the morph shows **fusion** or **cumulation** of two separate meanings.
2. *One meaning, two morphs*

   One and the same function can be signaled (redundantly) by different morphs in a given word. A simple example is found in Latin, as shown in Table 11.5.

Table 11.5 Some active indicative forms of Latin AMO 'I love/I have loved'.

P/N (SG)	PRESENT	PERFECT
1st	am-o	am-a-v-i
2nd	am-a-s	am-a-v-isti
3rd	am-a-t	am-a-v-it

The *-v-* morph realizes perfect tense, and has no other function, so we can say that *-v-* is the **principal exponent** of (perfect) tense. However, the 1sg endings also differ with tense, and thus serve as **secondary exponents** of this category. This means that the meaning of 'perfect tense' extends over both *-v-* and *-i* in *amavi*. This is often referred to as **extended exponence**.

3. *One meaning, no morph – null morphemes*

   Notice that there is no ending in the genitive plural of Russian Class II and IV nouns in Table 11.1. In a morpheme-based theory we must say that this property set, 'genitive plural, ClassII/IV', is signaled (cumulatively) by a **null** or **zero morpheme**: *komnat-Ø*. Similarly, in derivation we often find cases of **conversion**, in which a word belonging basically to one category (such as the noun *chair* or the verb *run*) is used in another (the verb *to chair*, the noun *a run*). Given agglutination, this, again, would have to be handled by assuming a null morpheme.

4. *One morph, no meaning*

   A traditional type of meaningless morpheme is the famous **cranberry morph**. Words such as *blueberry, blackberry, cloudberry, cranberry* etc., are clearly compounds of *berry* and refer to types of berry, but what does "cran" mean? More subtly, we saw that the *black* of *blackbird* doesn't have any meaning, strictly speaking. Aronoff (1976) argues in detail for English that there are cranberry morphs which have morphological properties (show allomorphy) and which therefore have to be regarded as morphemes. Thus, a verb such as *understand* is derived morphologically from the prefix *under-* (as in *underwrite, undertake, undermine*, ...) and *stand* (as in *withstand*). This is clear because they have the same irregularity in the past tense as the base verb (*understood*, *withstood*). However, neither the prefix nor the base preserves its meaning, or any meaning. I return to such cases in Section 6.2.

Cranberries are the examples of meaningless morphs most often cited, but the phenomenon is actually more widespread and more subtle. Thus, the adjectives in (20) illustrate a case in which a morpheme can be said to be meaningful only by stretching the meaning of "meaning" rather uncomfortably:

(20)

Noun	**Adjective**	
morphology	morphological	morphological theory
navy	naval	naval uniform
poetry	poetic	poetic license
nerve	nervous	nervous system

These are different from normal adjectives ending in the same suffixes such as *topical*, *sympathetic,* or *adventurous* in that they don't express qualities or properties. Thus, we can say *very topical article*, *unsympathetic remark*, *highly adventurous project* but the adjectives in (20) can't be modified in this way: **very morphological theory*. The reason is that the adjectives in (20) are really no different from the basic nouns but used in the syntactic contexts where an adjective is needed, i.e., to modify a noun. Indeed, in a number of cases we can idiomatically replace such phrases with compounds: *morphology theory*, *navy uniform*, or, marginally, *nerve system*. Thus, the

derivational morphology which creates the adjectives changes the syntactic category of the word but doesn't add any element of meaning and thus, strictly speaking, is a kind of cranberry suffix. This type of category-shifting morphology is often referred to as **transposition**.

3 Morpheme Order

The order of morphemes in a word is usually strictly determined, even in languages with very free word order, and linguistic theory has to have some set of mechanisms for guaranteeing this order. A simple example of polymorphemic inflection is provided by nouns in Finnish, a typical "agglutinative" language. It has number, case, and possessor inflection on nouns, with a separate formative for each function. Some examples of inflected forms of the word *talo* 'house' are given in (21) (data from Karlsson 1987):

(21)	talo-ssa-ni	'in my house'
	talo-lle-mme	'onto our house'
	talo-i-sta-si	'out of thy house'
	talo-i-lta-nne	'off of your houses'

On a morphemic, word syntax account, a word form such as *taloissani* 'in my houses' might have the form seen in Figure 11.2.

How do we ensure that the morphemes come in this order and not, say, **talossaini*, or **talonissai*?

A very influential approach to this problem is **Distributed Morphology** (Halle and Marantz 1993). This approach denies the distinction between morphology and syntax: all word structure is really an extension of syntactic structure and the affixes are morphemes compounded with each other by syntactic processes. In practice it is impossible to discuss the proposals of Distributed Morphology without also discussing Minimalist syntactic theory, which would take us well beyond the scope of this introduction. A handbook introduction to the model is found in Embick and Noyer (2007) and the papers collected in Matushansky and Marantz (2013) provide a sample of some of the issues that the model addresses.

More general problems of affix ordering are discussed in Manova and Aronoff (2010) and Rice (2011).

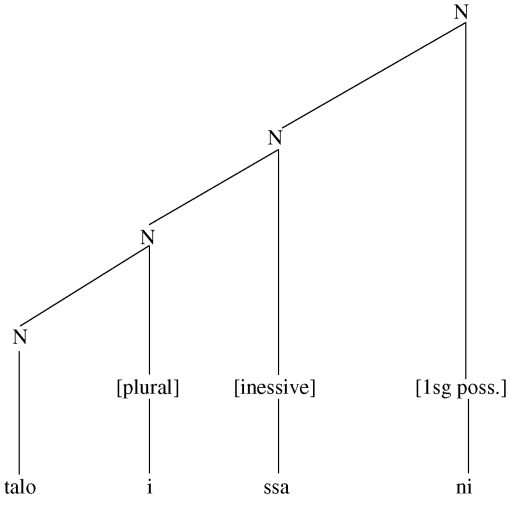


Figure 11.2 Word syntax tree for Finnish *taloissani* 'in our houses'.

4 Rule Function Morphology

The morpheme concept thus serves even agglutinative languages like Finnish rather badly. An alternative conception has been argued for by many morphologists (see Anderson 1992; Aronoff 1994; Stump 2001; Stewart and Stump 2007), under which affixation is just one of a set of phonological operations which can be performed on a base, triggered by the inflectional properties of the whole word. This is a variant of the classical **Item-and-Process** approach to morphology but I shall refer to it as the **Rule Function** approach (since processes are stated as rules which are usually thought of as something like functions in the mathematical sense).

To handle Finnish nouns we would set up a battery of rules applying in three blocks, one essentially for number marking, one for case, and one for possession. We start with a complete inflectional characterization of the word, say, {plural, inessive, 1sg.possessed}. This triggers three rules which take the current stem and add the appropriate suffixes:

(22) a. {plural}(X) = X+i
b. {inessive}(X) = X+ssa
c. {1sg.possessed}(X) = X+ni

A rule such as {plural}(X) is intended to be read "the plural form of (the base form of) X is constructed by adding *-i* to X." Taking the base *talo* these build up the required form in the obvious way. Note that there is no need for subcategorization since morpheme order is reflected in the ordering of the rule functions. The "zero affixes" of the Singular and the Nominative are handled very naturally: there is no rule corresponding to these properties, therefore nothing is done to the base form. There is thus no need for dubious constructs such as strings of null morphemes.

The nominative plural form is defined by a rule similar to (22) but which is more specific, shown here in (23):

(23) {plural, nominative}(X) = X+t

Now, where we wish to realize the nominative plural form of a lexeme both rule (22a) and rule (23) could apply, so we need a way of ensuring that it is (23) that applies, not (22). The rules (22a, 23) are regulated by a very general principle (often called the **Elsewhere Condition**): if two rules can apply to the same base, it is the more specific which wins out, in this case (23). The more general of two rules defines the **default** case, and this default is pre-empted or overridden by the more specific rule. The use of the notion of default in morphology has become extremely important in recent research and some form of it is even accepted by protagonists of (somewhat modified) types of Item-and-Arrangement theory (Halle and Marantz 1993; Wunderlich 1996).

The rules of (22) are affixations but could just as easily be any phonological process, such as vowel ablaut or reduplication. The Rule Function approach rejects the idea that affixes are lexemes compounded with base lexemes. Rule functions can also handle syncretisms very neatly. These can be a problem for morpheme-based theories because they are often defined over parts of the paradigm independently of the actual affixes. A simple example of this is the relationship between the perfect participle (*has cooked*) and the passive participle (*was cooked*). These are identical for all verbs despite variation in form (*has/was written*, *has/was sung*, etc.). It is hard to see how the generalization can be stated over such diverse "lexemes" as *-ed*, *-en*, *u*-ablaut, and so on. This can be captured in the morphology by stating a **rule of referral** for constructing passive participles from perfect participles as in (24):

(24) {Passive Participle}(X) = {Perfect Participle}(X)

As Stump (2001) argues, (24) is just another type of rule function and can interact with other rule functions in a variety of ways. To achieve the same effect, the morpheme-based Distributed

Morphology model proposes rules of Impoverishment, which delete some of the feature content of morphemes and then redefine that content in terms of default values. Baerman, Brown, and Corbett (2005) argue that such a mechanism is unable to account for all the attested types of syncretism, however.

5 Paradigms and Principal Parts

The notion of **paradigm** plays an important role in rule function models. The important version of this notion is the set of grammatical words, that is, the word forms of a lexeme together with their grammatical description, usually defined in terms of a feature system. In traditional grammars of inflectionally complex languages with arbitrary inflectional classes such as Latin a useful pedagogical tool is the **principal part**. This is a specific grammatical word or set of grammatical words which allows a learner to figure out all the other word forms in the paradigm of that lexeme. For instance, Latin primers provide four verb forms from which the rest of a verb's paradigm can be predicted, the 1sg active present indicative, the infinitive, the active present perfect, and the passive perfect participle. Recent research has shown that there are rich and interesting dependencies between parts of paradigms which go well beyond this traditional notion of principal part. Specifically, for any word form it is possible to calculate the probability of correctly predicting any other word form of that lexeme. This probability can be calculated as a measure known as **entropy** (derived from statistical physics and thermodynamics via telephone engineering). This is a measure of how predictable a form is given another form in the paradigm. Ackerman and Malouf (2013) argue that there is a (fairly low) limit to the entropy values found in natural language systems. Stump and Finkel (2013) provide a very detailed (but rather technical) summary of recent research and its implications.

6 Lexeme Structure and Lexical Relatedness

6.1 *Derivational morphology*

We saw in Section 1.2 that compounds show varying degrees of semantic transparency: *morphology article*, *blackbird*, *husband*. Much the same can be said of derivational morphology. In (25) we see examples of fully transparent, compositional derivation:

(25) a. cat-like
b. elephant-like
c. lion-like
d. ape-like

These all mean roughly "like a typical X" where X = {cat, elephant, lion, ape, ...}, and we can call such forms **similitudinal adjectives**. This derivation is highly productive, in that "X" can be virtually anything: speakers can understand and use a coinage like *iguana-like* without ever having heard it before (though the exact force of *-like* is rather subtle, e.g., what is the precise semantics of *an iguana-like expression/skin/gait*?). This type of suffixation is so transparent that it resembles compounding – indeed, some might claim that *-like* adjectives are in fact compounds (though of a very rare type in English). Now contrast (25) with the examples in (26):

(26)	a.	cat	catty	cat + y
	b.	elephant	elephantine	elephant + ine
	c.	monster	monstrous	monst(e)r + ous

Although these may also be similitudinal adjectives, they are not compositionally derived from their bases. This means that the base has lost all meaning and functions much like the cranberry morpheme *stand* in *understand*.

Much derivational morphology is similar to that of (26), if not quite so dramatically so. For instance, *hopeless* means "without hope," but this isn't true for all uses. If we say "I wouldn't pick him for the job, he's hopeless" we are saying that he is extremely bad and unsuitable. He himself is not necessarily without hope in the literal sense (he could be one of life's irrepressible optimists). Such cases are the norm and it turns out that there is a cline of transparency running from *cat-like* to *hopeless* to *catty* to *understand*.

6.2 Four types of lexical relatedness

Transparent derivational morphology defines a network of relatedness amongst lexemes, but it is only one of four types of relatedness, one in which morphological relatedness goes hand in hand with semantic relatedness. The second type of relatedness is that mediated solely through semantics, without any morphological relatedness. For example, there are similitudinal adjectives which mean "like *Noun*" but which aren't morphologically derived from *Noun* e.g., *infantile* (≠ "like an infant"!), or *puerile*, both meaning "child-like" (with additional pejorative overtones). These could be said to end in an affix *-ile*. I cited *writer* in Section 2.1 as an example of a subject nominalization, and this represents a very productive formation, but not all verbs permit it. The subject nominal corresponding to the verb *fly*, as in "fly a plane (professionally)," is an entirely different word, *pilot*. Admittedly, the form *flier* exists for other uses of the verb, but we wouldn't say, for instance, **Tom was the flier of that Boeing 747.* This is a kind of derivational suppletion.

The third type of relatedness is that represented by **systematic polysemy**, that is, where we have two different lexemes with the same form. A familiar example is the alternation shown in (27) (referred to variously as **inchoative**, **causative**, or **anti-causative**). Most linguists would probably say that there are two distinct, though related, BREAK lexemes here:

(27) a. Tom broke the vase
b. The vase broke.

In many languages such usages are conveyed morphologically (by what is usually considered derivational morphology). Notice that the verb retains all its purely morphological properties in both usages, so there is no conversion or affixless derivation in the normal sense here. Rather, we seem to have two closely related lexemes which share all the same word forms.

In subsection 2.2.3 I pointed out that both the prefix and base of *understand* are cranberries. This is the fourth type of derivation in which there is clear morphological relatedness but no semantic connection (*asemantic relatedness*). In Russian this phenomenon is pervasive. Most Russian verbs are derived by prefixation of a few hundred simplex verb roots. In some cases the prefix and root contribute to the overall meaning compositionally, but in many cases it is impossible to ascribe a clear meaning to either root or prefix, just as with *understand*. Thus, from the verb *p´isat´* 'write' we obtain *pr´i-p´isat´* 'ascribe', *o-p´isat´* 'describe', *pod-p´isat´* 'sign', and so on. Each of the prefixes occurs in hundreds or thousands of verbs (sometimes with identifiable meaning). Moreover, each of these formations belongs to the same special subclass of Class I that *p´isat´* belongs to (thus, the present tense stem is *p´iš-*, rather than the default stem form for Class I, **p´isaj-*,) and they all show the same pattern of stress shift as the base verb. Finally, they all behave like morphologically prefixed forms. We can see this because the base verb, *p´isat´*, is imperfective in aspect, but nearly all prefixed verbs in Russian are perfective and form a special secondary imperfective (usually by suffixation of *-iv-*). Sure enough, all the derived verbs from *p´isat´* are perfective and form their secondary imperfective in *-iv-*: *pr´i-p´is-iv-at´*, *o-p´is-iv-at´*, *pod-p´is-iv-at´*, etc. Russian verbs thus present a much more convincing demonstration of Aronoff's original

point because there can be no doubt that the verbs are prefixed, and because the base exists in nearly all cases as an independent verb with exactly the same morphological properties, the majority of the native verb lexemes in the language are like this (indeed, the majority of monomorphemic, nonprefixed verb stems are loans).

This property of the Russian lexicon is particularly damaging to the classical morpheme theory. Of the 28,500 or so verbs in Zaliznjak's Grammatical Dictionary of Russian, roughly 24,000 are prefixed. Of these a large proportion are highly regular aspectual or Aktionsart formations (which could be claimed to be more like inflections than lexeme-forming derivations). Given this, I would estimate that between a third and a half of these 24,000 are like the derivates of *p´isat´* discussed above. Thus, if we consider cases which are uncontroversially independent lexemes it turns out that the *majority* of Russian verbs consist of a cranberry prefix and a cranberry root. The significance of such cases has been significantly underplayed in the literature.

One might wish to claim that there is a fifth type of relatedness illustrated by denominal verbs in English formed by conversion (though deverbal nouns, such as *a bite* or *a broadcast*, would do as well). Verbs such as *to saddle (a horse), to shelve (books), to skin (a rabbit), to paper (a wall)*, and many others are clearly derived from nouns, but without any overt morphology. Presumably we would wish to say that this created new lexemes (it seems far-fetched to regard *to saddle* as an inflected form of the base noun) and hence constitutes a derivational relation. However, this can either be regarded as a type of derivation which happens not to involve morphology (and hence a subtype of standard derivation) or a subtype of systematic polysemy.

In sum: lexemes can be related to each other by (i) morphology which induces a compositional meaning change; (ii) systematic meaning relation which is not matched by any formal relatedness (*suppletive derivation*); (iii) systematic meaning relation between different meanings associated with the same form (*systematic polysemy*); and (iv) purely in terms of shape, *asemantic relatedness*. These extremes define a space within which word relatedness can vary, so that *catty* could be said to be an example of derivation with respect to the suffix but asemantic relatedness with respect to the base. A recent summary of these issues is found in Spencer (2013).

6.3 Mixed categories

We end this section by looking at a set of cases which occupy a borderline position in some sense, and which are currently the focus of a good deal of research effort. A very common form of verb-to-adjective transposition is illustrated by **participles**. These are adjectival forms associated with verb lexemes, often expressing verbal tenses, aspects, or voices, but not adding any lexical meanings and hence usually considered inflectional forms. Examples in English would be the present and past/passive participles of *running water* or *a snatched/stolen kiss*. In many languages it is particularly obvious that the participles are adjectives, for instance, they not only modify nouns but also agree with them in number, gender, or case (something verbs don't normally do).

Participles illustrate an intriguing problem, as shown by the Russian examples in (28):

(28) a. Devuška čitaet gazetu
girl.NOM reads newspaper.ACC
'The girl is reading the newpaper'

b. devuška čitajuščaja gazetu
girl.NOM reading.F.NOM.SG newspaper.ACC
'the girl reading the paper'

In (28a) we see that the transitive verb 'read' takes a direct object in the accusative case. The participle in (28b) takes the same direct object marked in the same way but corresponding to

the subject in (28a) is the noun modified by the participle, 'girl', with which the participle agrees just like an adjective would (cf. (13)). This shows that the participle is not like a normal adjective because Russian adjectives do not take complements (especially not in the accusative case!). Participles, however, take exactly the same set of complements as their base verb, and mark them in exactly the same way as the verb. Forms with this Janus-like behavior are often referred to as **mixed categories**. Deverbal nominalizations provide further instances. Thus, in *Tom's writing the letter would be surprising* the nominal *writing* expresses an object in the manner of a verb (*Tom wrote the letter*) not a noun (cf. *Tom's writing of the letter*) but expresses its subject in the manner of a noun (cf. *Tom's letter*). A detailed overview of the phenomenon of transposition and its implications can be found in Spencer (2013).

6.4 Complex predicates

We began the chapter with a discussion of lexical integrity and a good deal of recent research has been devoted to clarifying this notion, and hence the notion of "word." Equally, considerable research effort has been directed in recent years to cases in which there is a mismatch between the number of lexemes and the number of syntactically realized word forms. Such constructions are often referred to as **complex predicates**, a term which is used broadly of two types of phenomenon: (i) a single phonological, syntactic word form corresponds to two lexemes and (ii) two phonological, syntactic word forms correspond to one lexeme. We saw one example of the type (i) complex predicate when we discussed noun incorporation (see example (5), Section 1.2). In this subsection I briefly mention two type (ii) cases.

A simple example of a type (ii) complex predicate is provided by an English phrasal verb such as *turn... off*. In *Tom turned the light off* we have a single verb lexeme *turn off* with the meaning "extinguish," but the two components can be separated by the verb's object. In the general case, we cannot predict the meaning of the phrasal verb from its components (compare for instance *Low temperatures will slow the process up/down* or *They ran a huge bill up*). A similar phenomenon is found in Hungarian, but with preposed particles ("preverbs"). Thus, *megérkez-* 'arrive' has the preverb *meg-*. In (29) we see the preverb (PV) functioning as a prefix to the verb (the prefix receives the initial word stress, for instance; the accent in Hungarian orthography indicates vowel length, not stress):

(29) Meg-érkezett
MEG-arrived
'S/he arrived'

However, in certain morphosyntactic circumstances (negation, questions, focusing) it can appear separated to the right of the verb, as shown in (30). (See Ackerman and LeSourd 1997 for discussion of such constructions).

(30) Nem érkezett meg
NEG arrived MEG
'S/he didn't arrive'

In (31) we see forms of the verb *meg-néz-* 'to watch, look at' as the complement of the verb *akar-* 'want', where it remains prefixed to the verb:

(31) Nem akarom meg-nézni ezt a filmet
NEG I.want MEG-watch this the film
'I don't want to watch this film'

However, when the main clause is neither interrogative nor negative, as in (32), we find that the preverb appears to the left of the main verb:

(32) Meg akarom nézni ezt a filmet
MEG I.want watch this the film
'I want to watch this film'

There is good reason to regard such phrasal verbs as single lexical items, i.e., lexemes: the meaning is often (though not always) idiosyncratic and, in Hungarian, processes which derive nouns or adjectives from verbs often apply equally well to the phrasal verbs. This is illustrated below, where a simple verb (33) is compared with a particle verb (34) (Ackerman and LeSourd 1997: 89):

(33) a. old-ani
'dissolve'
b. old-ás
'(chemical) solution'
c. old-ható anyag
'dissolvable substance'
d. old-hatatlan anyag
'insoluble substance'

(34) a. meg-old-ani
'solve'
b. meg-old-ás
'solution (to problem)'
c. meg-old-ható feladat
'solvable task'
d. meg-old-hatatlan feladat
'unsolvable task'

Thus, in Hungarian, a single lexeme, *meg-oldani*, can be systematically realized as more than one word in the syntax. Ackerman and LeSourd argue that this calls for a more sophisticated concept of lexical integrity: word forms such as 'meg', 'oldani', 'turn', and 'off' are single indivisible words which cannot be split up once they appear in sentences, and thus they exhibit lexical integrity. However, a given lexeme may be realized by a combination of such words, (*meg=oldani*, *turn=off*), and these may be separated in the syntax, so that, as lexemes, they do not exhibit lexical integrity. In other words, lexical integrity is a property of word forms but not necessarily of lexemes.

7 Conclusions

The notion "word" covers several distinct linguistic concepts, including: lexeme, word form, grammatical word. Not all the properties of words can be explained in terms of syntax or phonology; in particular, the existence of arbitrary inflectional classes demonstrates the need to treat morphology as an autonomous component of grammar. The classical sign-based concept of the morpheme has been extremely influential in thinking about the internal structure of words, but this has been largely abandoned, at least for inflection, where morphologists increasingly appeal to the notion of rule functions and defaults to capture the structure of paradigms and the order of elements, and to account for deviations from the "ideal" of agglutinating morphology. In addition, the notion of principal parts in conjunction with the entropy measure is proving valuable in analyzing the structure of paradigms.

We surveyed four types of derivational relatedness, showing that words can be related to each other in four main ways: in terms solely of semantics, with no morphological relationship; in terms purely of morphology, with no semantic relationship; in terms of polysemy, in which there is a semantic relationship but the word forms remain the same; and the standard case, in terms of a semantic relationship mediated by morphology. We also looked at important cases of mismatch between form and function, the mixed categories, and complex predicates.

SUGGESTIONS FOR FURTHER READING

Aronoff, M. and Fudeman, K. (2005). *What is Morphology?* Oxford: Blackwell Publishers.

Bauer, L., Lieber, R., and Plag, I. (2013). *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.

Booij, G. (2012). *The Grammar of Words*, 3rd edn. Oxford: Oxford University Press.

Haspelmath, M. and Simms, A. D. (2010). *Understanding Morphology*, 2nd edn. London: Hodder Education.

Katamba, F. and Stonham, J. (2006). *Morphology*. London: Palgrave Macmillan.

NOTES

1 Parts of this chapter are based on work conducted as part of research funded by the Economic and Social Research Council (Project Reference R000236115), to whom I express my gratitude. I am grateful to Mark Aronoff for helpful suggestions for improvement.

REFERENCES

Ackerman, F. and LeSourd, P. (1997). Toward a lexical representation of phrasal predicates. In A. Alsina, J. Bresnan, and P. Sells (eds), *Complex Predicates* (pp. 67–106). Stanford, CA: CSLI Publications.

Ackerman, F. and Malouf, R. (2013). Morphological organization: the low entropy conjecture. *Language*, 89, 429–64.

Anderson, S. R. (1992). *A-morphous Morphology*. Cambridge: Cambridge University Press.

Aronoff, M. (1976). *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Aronoff, M. (1994). *Morphology by Itself*. Cambridge, MA: MIT Press.

Baerman, M., Brown, D., and Corbett, G. G. (2005). *The Syntax-Morphology Interface: A Study of Syncretism*. Cambridge: Cambridge University Press.

Bauer, L. (2001). *Morphological Productivity*. Cambridge: Cambridge University Press.

Bauer, L. (2009). Typology of compounds. In R. Lieber and P. Štekauer (eds), *The Oxford Handbook of Compounding* (pp. 343–56). Oxford: Oxford University Press.

Booij, G. (2012). *The Grammar of Words*, 3rd edn. Oxford: Oxford University Press.

Brown, D., Chumakina, M., Corbett, G. G., Popova, G. D., and Spencer, A. (2012). Defining 'periphrasis': key notions. *Morphology*, 22, 233–75.

Corbett, G. G. (2010). Canonical derivational morphology. *Word Structure*, 3, 141–55.

Corbett, G. G. and Fraser, N. (1993). Network morphology. *Journal of Linguistics*, 29, 113–42.

Embick, D. and Noyer, R. (2007). Distributed Morphology and the syntax-morphology interface. In G. Ramchand and C. Reiss (eds), *The Oxford Handbook of Linguistic Interfaces* (pp. 239–88). Oxford: Oxford University Press

Halle, M. and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale and S. J. Keyser (eds), *The View from Building 20* (pp. 111–76). Cambridge, MA: MIT Press.

Karlsson, F. (1987). *Finnish Grammar*. Helsinki: Werner Sörderström Osakeyhtiö.

McCarthy, J. J. (1982). *Formal Problems in Semitic Phonology and Morphology*. New York: Garland Press.

Manova, S. and Aronoff, M. (2010). Modelling affix order. *Morphology*, 20, 109–31. [Special Issue: *Affix Combinations*, edited by Stela Manova and Mark Aronoff].

Matushansky, O. and Marantz, A. (eds) (2013). *Distributed Morphology Today. Morphemes for Morris Halle*. Cambridge, MA: MIT Press.

Mithun, M. (1984). The evolution of noun incorporation. *Language*, 60, 847–94.

Muravyova, I. A. (1998). Chukchee (Paleo-Siberian). In A. Spencer and A. Zwicky (eds), *Handbook of Morphology* (pp. 521–38). Oxford: Blackwell Publishers.

Ramchand, G. and Reiss, C. (eds) (2007). *The Oxford Handbook of Linguistic Interfaces*. Oxford: Oxford University Press.

Rice, K. (2011). Principles of affix ordering: an overview. *Word Structure*, 4, 169–200.

Skorik, P. Ja. (1977). *Grammatika ⊠ukotskogo jazyka, tom 2*. [*A Grammar of Chukchi, vol. 2*. In Russian]. Leningrad: Nauka.

Spencer, A. (2013). *Lexical Relatedness: A Paradigm-Based Model*. Oxford: Oxford University Press.

Spencer, A. and Luís, A. (2012). *Clitics: An Introduction*. Cambridge: Cambridge University Press.

Spencer, A. and Zwicky, A. (eds) (1998). *Handbook of Morphology*. Oxford: Blackwell Publishers.

Stewart, T. and Stump, G. T. (2007). Paradigm Function Morphology and the morphology-syntax interface. In G. Ramchand and C. Reiss (eds), *The Oxford Handbook of Linguistic Interfaces* (pp. 383–421). Oxford: Oxford University Press.

Stump, G. T. (2001). *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Stump, G. T. and Finkel, R. A. (2013). *Morphological Typology: From Word to Paradigm*. Cambridge: Cambridge University Press.

Toman, J. (1998). Word syntax. In A. Spencer and A. Zwicky (eds), *Handbook of Morphology* (pp. 306–21). Oxford: Blackwell Publishers.

Wunderlich, D. (1996). Minimalist morphology: the role of paradigms. In G. Booij and J. van Marle (eds), *Yearbook of Morphology 1995* (pp. 93–114). Dordrecht: Kluwer.

12 The Lexicon

D. A. CRUSE

1 Introduction

To take a simplistic view, the bare essentials for a language are, first, a set of basic units, and second, a set of rules for combining them into larger, more complex units like phrases and sentences. A list of the basic units constitutes the lexicon of the language; a specification of the combinatory rules constitutes the grammar.

The basic units must have both a form and a meaning (in the broadest sense); the entries in the lexicon must specify these, together with information necessary for the proper application of the grammatical rules. The combinatory rules will tell us not only what complex forms are allowed in the language, but also how their meanings are to be computed.

What are the units that are listed in the lexicon? The obvious answer is that they are words, and that is what we shall take them to be (although the matter is not quite so straightforward as it might at first seem). To the layperson, probably the most important thing about a word is what it means; this chapter has a similar bias, being chiefly about words and their meanings. We begin by looking at what sort of things words are, as a linguist sees them.

2 Words

It is notoriously difficult to frame a definition of a word which is satisfactory for all languages, and even for everything word-like in a particular language. We shall assume that, as in Wittgenstein's famous example of *game*, no succinct definition applicable to all cases is possible, and that the best approach is to look for features characteristic of central examples of the class.

2.1 Lexical forms, lexical units, and lexemes

The word *word* is used in different senses, and it will be as well to clarify the most important of these right from the start. Suppose we are doing a crossword puzzle. It is quite possible that, say, *walk* is the correct answer to a clue, but *walked* is not: from this perspective *walk* and *walked* are different words. Now imagine someone who encounters sentence (1):

(1) I have disconfirmed the doctor's prognosis

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

and asks: "Is there such a word as *disconfirm* in English? Look it up in the dictionary." What does *word* mean here? Clearly not what it means in the crossword context, since the dubious sentence contained *disconfirmed*, but the question asks about *disconfirm*. Let us at this point make a terminological distinction between *word forms* and *lexemes*. We shall say that *walk* and *walked*, and *disconfirm* and *disconfirmed* are different word forms, but whereas *walk* and *disconfirm* represent different lexemes, *walk* and *walked* are different word forms belonging to the same lexeme.

What, then, are lexemes? As a first step let us say that they are the units listed in a dictionary. A dictionary provides a list of the lexemes of a language, each indexed by one of its word forms. (Which word form a dictionary uses to indicate a lexeme is at least partly a matter of convention. For instance, in English, for verbs, we use the bare stem: *run*, *walk*; in French it is the infinitive: *courir*, *marcher*; in Latin, the first person singular of the present indicative: *curro*, *ambulo*.)

A more technical characterization is that a lexeme is a set of related meanings associated with a set of related word forms. Sometimes meanings associated with a single word form are clearly unrelated, as in the case of *bank* (financial) and *bank* (river); these would therefore be assigned to different lexemes. In other cases a relationship can easily be intuited, as with *position* (location), *position* (opinion), and *position* (job), and these will be considered to belong to the same lexeme. Most dictionaries give separate main entries to distinct lexemes, even if they share the same forms, but group related meanings under a single main entry. What we shall call "a set of related word forms" is a set of forms which differ only in respect of inflectional affixes (such as the singular and plural forms of nouns, or the past, present, and future forms of verbs).

This is fine, but how do we then designate the three individual items *position*? It is usual to call the distinct meanings *senses*, but what is the sound-meaning complex? I shall call them *lexical units*. Actually, in many (perhaps most) contexts it is perfectly clear what *word* means: the expressions *word form*, *lexeme*, and *lexical unit* will therefore only be used when there is a danger of confusion.

2.2 *Individuating word forms: graphic and phonetic clues*

Most modern writing systems (English is no exception) indicate word (here, obviously, "word form") boundaries by means of spaces. This makes reading a lot easier. However, there is usually no analog of written spaces in spoken language, although this usually comes as a surprise to the layperson, because spoken words are clearly demarcated perceptually. There may, nonetheless, be signals of other types which indicate the positions of the boundaries of spoken words. For instance, many languages have a regular stress pattern for words, as in Czech, where words are always stressed on the first syllable. Other signs may be more complex or subtle. For instance, to take a venerable example, English speakers can discern purely from the sound the different positions of the word boundaries in *night rate*, and *Nye trait*, and between *parks treat* and *Park Street*, at least when they are carefully pronounced, even though there is no silence between the words. This is because, for instance, the variety of /r/ which occurs at the beginning of a word is different from that which appears when it is preceded by /t/, and is different again if the /t/ is preceded by /s/.

2.3 *Grammatical properties of words*

Prototypically (we shall not explore the exceptions) the stretches of a sentence that constitute word(form)s can be recognized by the fact that they are the largest chunks of the sentence which cannot be interrupted by the insertion of new material. Take the sentence *The government is strongly opposed to denationalization*. The possibilities of inserting new material are as follows:

> The (present) government, (apparently), is (very) strongly (and implacably) opposed (not only) to (creeping) denationalization, but … etc.

It will be noticed that the insertions all respect word boundaries, and all sequences of two words can be interrupted.

The parts of a word also cannot be reordered (**ationizaldenation*), although, at least in languages with some freedom of word order, the words themselves can be rearranged (obviously to varying degrees).

2.4 *Semantic properties of words*

There are several constraints on what a word can possibly mean. First, though, a nonconstraint. It might be thought that there could not be a word meaning, for instance, "to eat corn flakes while standing facing south on a Sunday morning." However, a brief period of reflection should convince the reader that such a meaning is not really impossible, merely unlikely in our culture: in a society where corn flakes were ritually dedicated to the god of the south, it would not be at all surprising if such an action received a lexical label. We shall look at two more serious constraints on possible word meanings, conceptual unity and internal completeness.

Whatever attracts a lexical label must have some degree of conceptual coherence. Let us confine our attention to what can be referred to by a noun: in the broadest sense, these are "things." Prototypical things are characterized by spatial continuity and persistence through time. Nonprototypical things must have something which confers unity on them. In front of me as I write, I can see, among other things, a bottle of Buxton mineral water, a photograph of Commander Data from *Star Trek*, and a ball of string. Is there any chance that these could be designated collectively by a noun? In a sense, yes: they could constitute the whole of my worldly possessions, and there could be a name for this (on the lines of my "estate," when I die). But that would not be a name for that particular set of things. Alternatively, they could be the requisites for, say, the Klingon Ceremony of Nga (or whatever). But in the absence of some such extrinsic unifying factor, the items mentioned would not be (collectively) nameable.

What I am calling "internal completeness" is more easily illustrated than explained. Take the phrase *a very large man*. The notion that there should exist a word meaning "large man" is not at all exotic (think of *giant*); nor is the idea of a word meaning "very large" (e.g., *enormous*); there could well be a word meaning "very large man" (perhaps *colossus*), too. But what about a word meaning "very … man," i.e., a noun such that any adjective modifying it is automatically intensified? This, surely, offends against our deepest semantic intuitions: it is an impossible meaning for a word. The same would be true of a putative "word" *beve* meaning "drink chilled …," such that *beve wine* would mean "drink chilled wine" (words meaning "drink wine" or "chilled wine," or even "drink chilled wine" could not be ruled out). The explanation seems to be on the following lines. We first need to distinguish *dependent* and *independent* components of a semantic combination. The independent component is the one which governs the external relations of the combination as a whole. So, for instance, in *very large*, it is *large* which governs the combinability of the phrase *very large* with other items. Thus the oddness of, say, *?very large air* is due to a clash between *large* and *air* – there is no clash between *very* and *air* (think of *very cold air*). By similar reasoning, the independent item in *chilled wine* is *wine*, and in *drink chilled wine* is *drink*. This process of reasoning allows us to establish chains of (semantic) dependencies (it does not matter whether the elements in a chain are expressed as different words, or are incorporated into the meaning of a single word). For instance, the chain for *very large man* is:

"very" → "large" → "man"

and that for *drink chilled wine* is:

"chilled" → "wine" → "drink"

The constraint that we are looking at says that the elements that constitute the meaning of a word must form a continuous dependency chain, with no gaps needing to be filled by elements from outside the word.

3 Lexical Semantics

The study of the meanings of words within linguistics is called *lexical semantics*. Under this banner a variety of spheres of interest, theoretical orientations, and methods of study flourish.

3.1 Theoretical approaches

To a large extent, how one goes about the business of studying meaning depends on what picture one has of the sort of thing meaning is. Some grasp of the major options will be useful as a background to the more detailed discussions which follow. We shall concentrate on two issues, holism vs. localism, and the relation between linguistic meaning and concepts. Let us begin with the holism / localism debate. Essentially, a holist believes that the meaning of a word is fundamentally relational, that is to say, it is a matter of relations with other words in the language. A localist believes that a word's meaning is self-contained, and describable independently of the meanings of other words.

3.1.1 The contextual / holistic approach Within linguistics, what philosophers of language call *holistic* theories of meaning are usually called *contextual* theories. These come in several varieties: two will be briefly illustrated here.

The first type falls under the heading of *structural semantics*. The basic notion of the interdependence of meanings can be illustrated as follows. Think of a child learning the names of the animals. The fact that a child can say *It's a dog* every time s/he is given a dog to identify, does not prove s/he has grasped what *dog* means; just as important is that s/he should avoid saying *It's a dog* when faced with a cat, or fox, or whatever. In other words, the meaning of *dog* (or any other word) cannot be learnt in isolation. A structuralist such as Lyons (the seminal reference is Lyons (1963)) builds on this basic insight, and characterizes the meaning of a word as its position in a network of relationships. Let us consider what that would mean in the case of *dog*. First, *dog* belongs to a set of words with which it has an exclusion relationship, that is to say, *It's a dog* implies *It's not a cat / lion / camel / cow / squirrel / etc.*; furthermore, all these fall into the denotation of a more inclusive term *animal*. *Animal* (at least on one reading) also belongs to a set whose members are mutually exclusive (including *insect*, *fish*, *bird*, etc.); these in turn are included in *living thing*, and so on. But *dog* has many other relations, for instance, with *tail*, *paw*, *head*; with *pack*; with *bark*, *howl*; with *kennel*, which itself has relations with other structures such as *hut*, *cabin*, *house*, and so on. Ultimately, every word is linked, directly or indirectly, by means of specific links such as "is a," "is not a," "has a," "is part of," "lives in a," etc., with virtually every other word in the lexicon: on the holist view, the meaning of a word is not fully comprehended until all these links are known (although, obviously, some links are more central than others).

An alternative version of a contextual theory takes its origin from Wittgenstein's dictum: *Don't ask for the meaning, ask for the use*. This is suggestive, but lacking in precision as a basis for a theory of meaning: what, precisely, do we mean by *use*? J. R. Firth (quoted in Mackin (1978)) gave the notion a useful twist when he said: *Words shall be known by the company they keep*. This line of thinking was developed into a holistic theory of meaning by W. Haas. (Haas's ideas are not readily accessible in published form; a summary can be found in Cruse (1986: ch. 1).) Haas started out from the idea that every grammatically well-formed sequence of words was either fully normal semantically, like *The dog barked*, or to some degree abnormal, like *?The cat barked* or *?The dog evaporated*. He then argued that if two words differ in meaning, this fact will inevitably

be reflected in a difference of normality in some context or other. For instance, that there is a difference in meaning between *illness* and *disease* follows from the fact that *during my illness* is more normal than *?during my disease*. Haas went on to characterize the meaning of a word as its normality profile across all its grammatically well-formed contexts, actual or potential: absolute synonyms, on this view, are words which have the same normality in all contexts.

3.1.2 The componential / localistic approach A localist believes that the meaning of a word is a self-sufficient entity which in principle is finitely describable. Whereas holists tend to see the meaning of a word as a set of relations, either with other words, or with possible contexts, a localist will typically say that these relations are a consequence of the word's meaning.

The most popular varieties of localism portray the meaning of a word as a finite assemblage of elementary bits of meaning, each of which accounts for some aspect of the semantic behavior of the whole. These "semantic atoms" (variously known as *semantic components*, *semantic features*, *semantic markers*) are drawn from a finite inventory, and in the strongest versions of the theory are psychologically real (in the sense that if we knew enough about the brain we would be able to identify a distinctive neuronal structure corresponding to each feature), and they are universal (in the sense that they form part of the language capacity that each human being is born with). It is impossible to give a satisfactory picture of any of the existing systems in a short space, but the following examples will give the flavor of such analyses:

filly	=	[HORSE] [FEMALE] [YOUNG]
boy	=	[HUMAN] [MALE] [YOUNG]
kill	=	[CAUSE] [BECOME] [NOT] [ALIVE]
chair	=	[OBJECT] [FURNITURE] [FOR SITTING] [FOR ONE PERSON] [WITH BACK]

3.1.3 The conceptual approach Much debate centers on the relation, if any (but surely there must be some), between linguistic meaning and concepts, or, as far as we are concerned in this chapter, between word meanings and concepts. Earlier semanticists (including Lyons and Haas) did not believe that anything solid was known about concepts; they therefore preferred to pursue their semantic studies without reference to such entities. The rise of cognitive psychology has made concepts more respectable, and few would now deny their significance. The debate now concerns whether, or to what extent, meaning can be identified with concepts: do words map directly onto concepts, or is there an intermediate level of semantic structure where word meaning is to be located, and the connection with concepts indirect? The present author's sympathies lie with the conceptual approach. A conceptual (or "cognitive") semanticist would argue that there is no theoretical work for an autonomous linguistic semantic level to do that cannot be performed at the conceptual level. He would also argue that the connection between words and the outside world is mediated through concepts, and that therefore examining world–word relations is not the most profitable approach to word meaning.

4 How Many Meanings? Contextual Variability of Word Meaning

Most words are capable of presenting different semantic faces in different contexts. Sometimes the differences are major and clear cut, as in:

(2) The boat was moored to the bank.

(3) She works in a bank.

At other times the difference is more subtle:

(4) John's maths teacher is on maternity leave.
(5) Bill's maths teacher is on paternity leave.

Here we can infer from the context that John's maths teacher is a woman whereas Bill's is a man.

It is important to be able to decide whether two interpretations of a word in different contexts represent one semantic unit or two. This is not a purely theoretical concern: for instance, a lexicographer will have to decide how many definitions to give for *bank* and *teacher*. We shall take the position that the basic unit of word meaning is the *sense*, and we shall say that a word has X senses if and only if it is X-ways ambiguous. We now need to be more explicit about what it means for a word to be ambiguous.

4.1 Ambiguity

Consider sentence (6):

(6) We managed to get to the bank just in time.

In the absence of a biasing context, the two readings of *bank* are in competition with one another: like the two visual construals of a Necker cube, only one can be at the focus of attention at any given moment. In a particular context, a speaker will "intend" only one of the meanings and will expect the hearer to make the same selection. There is no general meaning of *bank* which subsumes the two alternatives, and the options of remaining uncommitted or of taking both meanings on board are not available (outside of deliberate word play). Contrast this with the following case:

(7) We shall talk to Mary's teachers.

Of course, the individual teachers referred to in (7) must be either male or female, but (a) the speaker may not even know the sex of the teachers involved and will not expect the hearer to select a particular gender; (b) there is a general meaning of *teacher* which covers both possibilities; (c) the sex of the teachers can be left unspecified; furthermore, sentence (7) may well refer to a mixed group. By the criteria suggested, then, *teacher* is not ambiguous, and does not have two senses corresponding to "male teacher" and "female teacher"; a lexicographer would not need to give two definitions for *teacher*.

Ambiguous words typically pass the traditional *ambiguity tests*.

4.1.1 The identity test In *John has changed his position; so has Mary* The word *position* must be interpreted the same way in both halves of the sentence: if John has changed his mind on some political issue, then that's what Mary did, too; likewise if John has changed his location. This shows that *position* is ambiguous. In contrast, in *I spoke to a teacher; so did Mary*, there is no pressure to interpret *teacher* in the same way (gender-wise) in each conjunct, hence *teacher* fails this test.

4.1.2 The independent truth-conditions test It is easy to think of a situation where one could truthfully answer the following question both in the negative and the affirmative:

(8) Have you had a drink in the last six hours?

This shows that the readings "take alcoholic beverage" and "imbibe liquid" are distinct senses. There is no comparable possibility for simultaneously true *Yes* and *No* answers to (9):

(9) Have you spoken to a teacher?

4.1.3 The zeugma test A context which activates more than one reading of an ambiguous word gives rise to a sense of punning:

(10) The old man expired on the same day as his driving license.

(11) When the chair became vacant, the University Appointments Committee sat on it for six months.

(The effect in (11) hinges on the ambiguity of both *chair* and *sat on*.)

4.2 Polysemy and homonymy

The alternative readings of an ambiguous word may be totally unrelated, as in the case of *bank*, or they may be related, as in the case of *position* (see below for some discussion of possible types of relatedness). An ambiguous word with unrelated readings is described as *homonymous*; if the readings are related, the word is said to be *polysemous*. Homonymous words are usually given two main entries in a dictionary; polysemous variants are normally listed under a single main heading.

5 Sense Relations

Sense relations are relations between word meanings. Of course, every word has a semantic relation of some kind with every other word, but not all such relations have any intrinsic interest. To be interesting, a relation must recur with significant frequency throughout the vocabulary, and must be capable of supporting significant generalizations. (A much fuller treatment of sense relations than can be accommodated here may be found in Cruse 1986.)

There are two major classes of sense relation, depending on the grammatical relation between the words bearing the senses, namely, *paradigmatic* and *syntagmatic* relations. Paradigmatic sense relations are relations between the meanings of words which can occupy the same syntactic slot, and serve to unite the range of lexical meanings available at a particular point in a sentence into a more or less coherent structure. Take, for instance, the (incomplete) sentence: *John grows a number of ___ in his garden*. There is a structured set of choices of words to fill the gap. One may choose very general words like *trees*, *flowers*, or *vegetables*, or something more specific, falling under one of the general terms, for instance, *conifers*, *cabbages*, *carnations*. We shall look at the structuring in such a set in more detail in a moment, but it can already be appreciated that the words provide an articulation of the experienced world.

Syntagmatic sense relations hold between words in the same phrase or sentence. Intuitively, some words "go together" semantically, while others "clash": consider *drink wine* and *drink water*, compared with *drink rock* or *drink sound*. There is a relation of cohesiveness between *drink* and *wine* which is absent from *drink* and *rock*. Syntagmatic sense relations are thus involved with the semantic coherence of grammatical strings.

5.1 Paradigmatic sense relations

It is paradigmatic relations which have received the most attention from linguists. For convenience, they may be divided into two sorts, relations of identity and inclusion, and relations of opposition and exclusion.

5.1.1 Relations of inclusion and identity I: hyponymy We begin with relations of inclusion. There are two basic types of these. In the first type, the inclusion is of one class in another, as in the case of *car* and *vehicle*, where cars constitute a subclass included in the larger class of vehicles; in the

second type, the inclusion is observable at the level of individual entities, as in the case of *finger* and *hand*, where every individual hand includes a number of fingers as parts.

The class-inclusion relation, called *hyponymy*, is exemplified by *dog:animal*, *apple:fruit*, *tulip:-flower*, *cathedral:building*, *beer:beverage*, *copper:metal*, *kitten:cat*, *mare:horse*, *actress:woman*, and so on; of the two related items the more specific is called the *hyponym* (e.g., *dog*, *apple*), and the more general is called the *superordinate* (less commonly, the *hyperonym*), e.g., *animal*, *fruit*. Notice that although *dog* is a hyponym of *animal*, it is a superordinate of, say, *spaniel*.

Hyponymy can be thought of as the "— is a —" relation which guarantees the truth of general statements such as *An apple is a fruit* and *An actress is a woman*. For a lexical item X to be a hyponym of another item Y, the truth of *An X is a Y* must follow logically from the meanings of X and Y. An expectation that if something is an X, it is likely to be also a Y, is not enough. For instance, if someone talks about a cat, most people will assume that the cat in question is somebody's pet. However, this does not entitle us to say that *cat* is a hyponym of *pet*, because there are cats which are not pets, and so *Cats are pets* is not automatically true by virtue of its meaning.

5.1.2 Relations of identity and inclusion II: meronymy The part–whole relation, in its lexical aspect, is called *meronymy* (sometimes *partonymy*); for instance, *finger* is a *meronym* of *hand*, and *hand* is the *immediate holonym* of *finger*. The notion of meronymy, like hyponymy, is relational rather than absolute: *hand*, for instance, is the holonym of *finger*, but it is at the same time a meronym of *arm*, which in turn is a meronym of *body*. The chain of relations stops at *body*, which may be termed the *global holonym*. Other examples of meronymy are as follows: *arm:body*, *petal:flower*, *engine:car*, *blade:knife*. Prototypical meronymous pairs (where X is a meronym of Y) are normal in frames such as: *X is a part of Y*; *A Y has an X*; *The parts of a Y are A, B, C … and so on*. Meronymy must be clearly distinguished from hyponymy, although both involve a species of inclusion. An easy way to highlight the difference is to note that a finger is not a kind of hand (meronymy), nor is a dog a part of an animal (hyponymy).

Not all portions of an object qualify as parts: a glass jug dropped on a stone floor does not break up into parts, but into pieces. The things we habitually call parts typically have a distinctive function or they are separated from sister parts by a formal discontinuity of some sort (or both). For instance, the wheels of a car have the function of allowing it to move smoothly over the ground, and transmit the motive power; the steering-wheel allows the direction of movement to be controlled; the door handles allow the doors to be opened and shut manually. Discontinuity manifests itself in a number of ways. For example, the wheels of a car are detachable and can move relative to the chassis; the fingers of a hand are not detachable, but have a certain freedom of movement; discontinuity may also be visual, like the cuff of a sleeve, or the iris of the eye.

Parts may be necessary or optional. The necessity in question is not a logical necessity, but a well-formedness condition: a hand with a finger missing is still a hand, but it is not a well-formed hand. In this sense, *finger* is a necessary (or *canonical*) part of *hand*, as is *prong* of *fork*. On the other hand, faces may be perfectly well formed without beards, and doors without handles – here we are dealing with optional (or *facultative*) parts. Some parts are more tightly integrated into their wholes than others. An indication of less than full integration is the possibility of describing the part as "attached to" its whole; this is typically not normal with fully integrated parts. Contrast *The handle is attached to the door* (not fully integrated) and *?The handle is attached to the spoon* (fully integrated).

5.1.3 Relations of identity and inclusion III: synonymy Dictionaries typically define synonyms on the lines of "words with the same or a similar meaning." This description undoubtedly applies to all words that we would intuitively call synonyms: *begin* and *commence*, *death* and *demise*, *wedding* and *marriage*, *motor* and *engine*. However, it is not restrictive enough, as it surely also applies to, for instance, *mare* and *stallion*, which both refer to horses, but which are not synonyms. It would seem useful, therefore, to examine more closely the notion of "same or similar meaning."

Synonym pairs or groups can be categorized according to how close the meanings of the words are. Three degrees of closeness can be recognized: *absolute synonymy*, *propositional synonymy*, and *near synonymy*.

The greatest possible resemblance between two senses is identity, in other words, absolute synonymy. A characterization of absolute synonyms based on Haas's contextual approach was offered earlier, namely, that they are equinormal in all (grammatically well-formed) contexts. This is based on the assumption that any difference of meaning will reveal itself as a difference in co-occurrence possibilities, hence the discovery of a context where one of the putative synonyms is more normal than the other rules out the pair as absolute synonyms. This is an extremely strict criterion, and a rigorous testing of candidate pairs leads rapidly to the conviction that absolute synonyms are hard to come by. From the semiotic point of view this should probably not be surprising: there is no obvious reason why a language should have two forms with absolutely identical meanings. Let us look at a few possible examples of absolute synonymy:

(i) nearly / almost:

These are shown to be not absolute synonyms by the differences in normality between (15) and (16), and between (17) and (18):

(15) We're very nearly home now.
(16) ?We're very almost home now.
(17) He looks almost Chinese.
(18) ?He looks nearly Chinese.

(ii) big / large:

The difference in normality between (19) and (20) is enough to disqualify these:

(19) You're making a big mistake.
(20) ?You're making a large mistake.

(iii) begin / commence:

These, too, are disqualified:

(21) Are you sitting comfortably, children? Then I'll begin.
(22) ?Are you sitting comfortably, children? Then I'll commence.

Absolute synonymy presumably approximates to what those people have in mind who maintain that true synonyms do not occur in natural languages.

There is perhaps a case for saying that absolute identity of meaning can occur between forms belonging to different varieties, especially dialects, of a language. An obvious example would be *fall* and *autumn* in American and British English, respectively. These are no different in principle to translational equivalents in different languages. Notice, however, that these would not come out as absolute synonyms by the Haasian test, since *fall* would be less normal than *autumn* in a sentential context that was otherwise lexically marked as British English. Saying that *fall* and *autumn* are identical in meaning presupposes a non-Haasian notion of what meaning is.

Propositional synonymy is less strict than absolute synonymy, and examples of this variety are consequently more numerous. It can be defined in logical terms: propositional synonyms can be substituted in any declarative sentence *salva veritate*, that is, without changing its truth value. By this criterion, *begin* and *commence* are propositional synonyms, because if *The lecture began at nine o'clock* is true, then so is *The lecture commenced at nine o'clock*, and vice versa.

There are too few absolute and propositional synonyms in any language to justify the existence of a dictionary of synonyms; the majority of what lexicographers call synonyms are, in our terms, near synonyms. The following illustrate sets of near synonyms:

(i) kill, murder, execute, assassinate
(ii) laugh, chuckle, giggle, guffaw, snigger, titter
(iii) walk, stroll, saunter, stride, amble
(iv) anxious, nervous, worried, apprehensive, fearful
(v) brave, courageous, plucky, bold, heroic
(vi) calm, placid, tranquil, peaceful, serene

The words in these sets are not necessarily propositionally identical, so for at least some pairs it is not anomalous to assert one member and simultaneously deny the other:

(23) He wasn't murdered, he was executed.
(24) They didn't chuckle, they tittered.
(25) He was plucky, but not heroic.

Near synonyms often occur normally in the test-frame *X, or rather Y*, which signals first, that Y conveys propositional information not present in X, and second, that the difference is relatively minor. Thus, (26) is normal, but (27) is odd, because the difference in meaning is too great; (28) is odd because there is no propositional difference:

(26) He was murdered, or rather, executed.
(27) ?He was murdered, or rather, beaten up.
(28) ?He was killed, or rather, deprived of life.

Near synonyms, then, are words which share a salient common core of meaning, but differ in relatively minor respects. There is at present no more precise characterization of "minor" in this context.

Synonyms (of all kinds) often occur in clusters, and it is common for the cluster to be centered round a neutral word which subsumes all the rest, and of which the others are a semantic elaboration. For instance, *kill*, *laugh*, *walk*, *anxious*, *brave*, and *calm* are the central items, respectively, in the sets detailed above.

5.1.4 Relations of opposition and exclusion I: incompatibility and co-meronymy We have looked at relations of inclusion; equally important are relations of exclusion, especially those that hold between sister items under a common inclusive term. Just as there are two sorts of inclusion, there are also two corresponding sorts of exclusion, which receive the labels *incompatibility* and *co-meronymy*.

Incompatibility is the relation which holds between, for instance, *cat* and *dog*, *apple* and *banana*, *rose* and *tulip*, *man* and *woman*, *church* and *supermarket*, *bus* and *tractor*. The essence of this relation is mutual exclusion of classes: if something is a cat, then it follows ineluctably that it is not a dog, and vice versa – there is nothing that is simultaneously a cat and a dog. The same is true for the members of the other pairs mentioned. Note that this is not simple difference of meaning. Take the case of *novel* and *paperback*, which are both hyponyms of *book*. They clearly do not mean the same; on the other hand, they are not incompatibles, because something can be simultaneously a novel and a paperback. The same applies to *mother* and *doctor*, and *tall* and *blonde*.

A parallel relation of exclusion applies to sister meronyms of the same holonym, as in *nose*, *cheek*, *chin* of *face*, or *wheel*, *engine*, *chassis* of *car*, and so on. Here the exclusion is (at least

prototypically) spatial: the sister parts of an individual whole do not have any material substance in common.

5.1.5 Relations of opposition and exclusion II: opposites Oppositeness and synonymy are the only sense relations likely to be familiar to a layperson. Most languages have an everyday word for opposites; the relation is cognitively very basic and quite young children can grasp the notion. Opposites are incompatibles of a special type: they are inherently binary, that is to say, they belong together naturally and logically in pairs.

Opposites fall into a number of relatively clearly defined types, the most important of which are *complementaries*, *antonyms*, *directional opposites*, and *converses*.

Complementaries are probably the most basic sort. They can be distinguished from noncomplementary incompatibles by the fact that negating either term logically implies the other. For instance, *Proposition P is not true* logically implies *Proposition P is false* and *Proposition P is not false* implies *Proposition P is true*; hence *true* and *false* are complementaries. They may be contrasted with ordinary incompatibles like *cat* and *dog*: *This is not a dog* does not imply *This is a cat*. Other complementary pairs are: *open:shut*, *dead:alive*, *stationary:moving*, *male:female*. A pair of complementaries bisects some conceptual domain, without allowing any "sitting on the fence"; whatever belongs in the domain must fall on one side of the divide or the other. (The negation test works only for items which belong in the domain presupposed by the test word: *This piece of chalk is not dead* does not imply *This piece of chalk is alive*, because chalk does not belong to the domain of things to which *dead* and *alive* properly apply.) The relation between complementaries can be portrayed as follows:

true	false

Antonyms (in the narrow sense – the term is also often used to refer to opposites in general) are gradable adjectives (i.e., ones which can be modified without oddness by intensifiers such as *very*, *rather*, *extremely*, *a little*, and so on). Typical examples are *long:short*, *fast:slow*, *heavy:light*, *difficult:easy*, *thick:thin*, *good:bad*, *hot:cold*, *clean:dirty*. They indicate degrees of some property such as speed, weight, or length, one term denoting a value on the scale above some implicit standard appropriate to the context, and the other term denoting a value lower than the standard. Unlike complementaries they do not exhaustively bisect their domain – there is a neutral area between them, which can be described as, for instance, *neither good nor bad*, *neither long nor short*, *neither hot nor cold*, etc. The relation between antonyms is portrayed in Figure 12.1.

The comparative forms of antonyms vary according to whether they presuppose the positive forms or not. For instance, for something to be *hotter* than something else, it has to be already *hot*: *X is cold, but it's hotter than Y* is therefore odd. On the other hand, something that is *longer* than something else does not need to be *long*. *Hotter* is known as a *committed* comparative; *longer* is *impartial*. Committedness can be used to define three classes of antonyms: *polar, overlapping*, and *equipollent*.

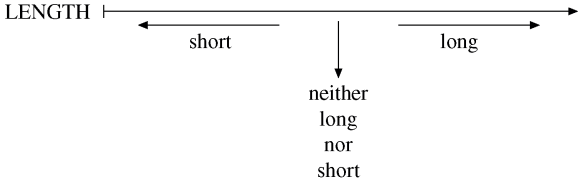


Figure 12.1 The relation between antonyms.

Polar antonyms: both members of a pair are impartial in the comparative:

(29) X is heavy, but it's lighter than Y.
(30) X is light, but it's heavier than Y.

Other examples are: *long:short*; *high:low*; *wide:narrow*; *thick:thin*; *fast:slow*; *hard:soft*. Polar antonyms indicate degrees of objective, usually measurable, properties.

Overlapping antonyms: one member of a pair is committed in the comparative, the other is impartial:

(31) ?X is good, but it's worse than Y.
(32) X is bad, but it's better than Y.

Other examples are: *kind:cruel*; *clean:dirty*; *polite:rude*. The members of this class all have an evaluative polarity, one member being commendatory, the other derogatory.

Equipollent antonyms: both members are committed in the comparative:

(33) ?X is hot, but it's colder than Y.
(34) ?X is cold, but it's hotter than Y.

Other examples are: *happy:sad*; *proud of:ashamed of*. The members of this group typically denote sensations or emotions.

Membership of one of the groups described above correlates with other properties, of which the following are worth noting. One important feature of antonyms is the possibility of degree questions. There are two principal forms, (a) those using a noun related to the adjective, as in *What is the length / weight / thickness of X?*, and (b) *how*-questions, such as *How long / thick / heavy is it?* The characteristics of the degree questions in each group are as follows:

(i) Polar antonyms: One antonym yields a neutral (impartial) *how*-question, the other (for most speakers) a somewhat abnormal question:

(35) How long is the piece of wood?
(normal and impartial)
(36) How short is the piece of wood?
(a bit odd, but if we have to interpret it, it is not impartial, but committed)

Polar antonyms also typically allow a *what*-question, but only with one of the terms of the opposition:

(37) What is the length of the piece of wood? (impartial)
(38) ?What is the shortness of the piece of wood?

For both types of degree-question, the term which produces an impartial question is the one which indicates "more of" the gradable property (e.g., *long* = "more than average length," *thick* = "more than average thickness," and so on).

(ii) Overlapping antonyms: One antonym yields a normal impartial *how*-question (e.g., *How good was the film?*) and its partner gives a normal, but committed *how*-question (e.g., *How bad were the exam results this year?*). In this case, it is the positively evaluative term which occurs in impartial questions, the other term being committed. Generally speaking, *what*-questions do not appear with antonyms from this group (*How clean was the room when you moved in? / ?What was the cleanness of the room when you moved in?; How polite was John when he came to see you? / ?What was John's politeness when he came to see you?*).

(iii) Equipollent antonyms: Normal *how*-questions are possible with both terms, but both are committed: *How cold was it?; How hot was it?* A *what*-question is possible for *hot:cold* (*What is its temperature?*), but this pair seems unusual in this respect.

An interesting property of overlapping antonyms is the feature of *inherentness*. Take the case of *bad:good*. Of two bad things, it is always possible to describe one as *worse* than the other: *The exam results this year were bad, but they were worse last year; This year's famine was worse than last year's*. However, the use of *better* is curiously restricted: *The exam results were bad this year, but they were better last year; ?This year's famine was better than last year's*. The general principle is that only things that are contingently bad (i.e., where good examples are conceivable) can be described using *better*: inherently bad things can only be qualified as *worse* (and, incidentally, cannot be questioned using *How good …?*: *?How good was John's accident?*).

Directional opposites: directional opposites are of two main types,

(i) static directions, like *up:down*, *backward:forward*, *north:south*:

West ◄——————————————► East

and

(ii) dynamic directional opposites (usually called *reversives*) such as *rise:fall*, *advance:retreat*, *increase:decrease*, *lengthen:shorten*, *dress:undress*, *tie:untie*, *mount:dismount*, *enter:leave*, *damage:-repair*, and so on.

We shall concentrate here on reversives. It will be noticed from the examples given that the notion of reversivity is extended from purely spatial domains to any sort of change of state. In general terms, a verb which is a member of a reversive pair denotes a change from an initial state (say S_1 to a final state S_2); its partner will then denote a change from S_2 to S_1, as shown in Figure 12.2.

An important feature of such verbs is that the path of change is irrelevant. For instance, a train entering then leaving a station may well travel in only one direction: what is important for entering is to start out "not in" something, and to end up "in" it, and the reverse is the case for leaving. Or take the case of tying and untying one's shoes: a film of someone untying their shoes is not identical to one of someone tying them run backward: the nature of the process of change is not specified by reversive verbs, only the initial and final states.

Reversive verbs have another curious property. Consider the following sentence: *Mary tied the knot, then untied it again five minutes later*. Assuming that *again* is unstressed, what is asserted to have been repeated, by the use of *again*? It is not, in fact, the act of untying – this may be the first time that Mary has ever untied a knot; what is said to recur is the state of being untied. This is presumably a further reflection of the importance of the initial and final states in the semantics of reversive verbs.

Converses: Converses are pairs like *above:below* and *parent:offspring*. Unlike most opposites, both terms can be used to describe the same state of affairs: for instance, *A is above B* means the same as *B is below A*, except in respect of which term serves as the reference point; similarly, *A is B's parent* designates the same relationship between A and B as *B is A's offspring*. For this reason, some linguists consider converses to be a variety of synonym.

Converses may be 2-, 3-, or 4-place relations, according to the number of arguments involved. *Above:below* are 2-place converses; *bequeath:inherit* are 3-place converses ("*John* bequeathed a

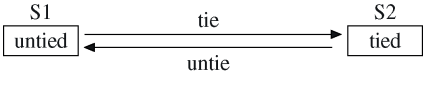


Figure 12.2 A reversive pair.

fortune to *Mary*" designates the same event as "*Mary* inherited a *fortune* from *John*"); *buy:sell* are 4-place ("*John* bought *the car* for *£1,000* from *Bill*" describes the same transaction as "*Bill* sold the *car* to *John* for *£1,000*").

5.2 Syntagmatic sense relations

We turn now to semantic relations between words which occur together in the same text. These can be roughly divided into two types. First, there are those which hold over relatively long stretches of text, between grammatically unrelated items, and which typically do not involve propositional meaning or directional properties. For instance, in *The Prime Minister attended the White House reception accompanied by her Dad*, there is a register clash between *Dad* and such formal items as *attended, reception* and *accompanied* (*Theresa May went to the White House party with her Dad* sounds less weird); the clash would be resolved by replacing *Dad* with the propositionally synonymous *father*. Notice that in this case, the clashing items are some distance from one another, and are not directly related grammatically (*Dad* does not clash with *her* nor *her Dad* with *by*). Second, there are relations which hold between closely related elements in the same grammatical construction and which do frequently involve propositional meaning and directional properties. For instance, the clashes in *?a male aunt, ?a highly strong man* and *?John drank a filing cabinet* involve the second type of relation. Take the case of the latter example: there is no clash between *John* and *drank,* or *John* and *filing cabinet*, the clash involves specifically *drank* and its direct object *filing cabinet*; the clash involves propositional meaning, since it can only be resolved by substituting either *drank* or *filing cabinet* with something propositionally different (e.g., *bought* or *wine*, respectively); *drank* imposes semantic restrictions on its direct objects (e.g., they must be liquids).

There are three possible effects of putting words together in a grammatically well-formed construction: either the result is normal, as in *John drank the wine*, or there is a semantic clash, as in *John drank the filing cabinet* or *a highly strong man*, or the result is pleonastic (or redundant) as in *a female aunt*.

Generally speaking, for a combination of words to be semantically normal, two conditions must be satisfied. If two words are joined together in a construction, it is usually possible to identify a selector, which imposes semantic conditions on possible partners, and a selectee, which satisfies (or does not satisfy) the conditions; the first requirement for a normal combination is that these conditions must be satisfied. This will avoid semantic clash. In the case of an adjective-noun combination, it is the adjective which is the selector: compare the ease with which semantic conditions (usually called selectional restrictions), can be specified for normality in the following:

A ___ woman.

There is no semantic generalization which covers, for instance, *intelligent*, *tall*, *pregnant*, *kind*, *highly paid*, *left-handed*, all of which combine normally with *woman*.

A pregnant ___.

Here, the restriction is much easier to capture: *pregnant* requires a head noun which (a) denotes a mammal, and (b) is not specifically marked as "not female" (like, for instance, *bull*).

It is generally the case that in a modifier-head construction, such as adjective-noun or intensifier-adjective, the modifier is the selector; in a head-complement construction, such as *drinks beer* in *John drinks beer*, it is the head of the construction, i.e., the verb, which is the selector.

In *?a female aunt* there is no semantic clash, but the combination is still odd, which indicates that a further condition must be satisfied. This is that a grammatically dependent item (modifier or complement) must contribute semantic information not already present in the head. Clearly, the notion FEMALE is part of the meaning of *aunt*, so the word *female* adds nothing,

and consequently *?a female aunt* is pleonastic, or redundant; *a lesbian aunt*, on the other hand, is semantically normal, because although *lesbian* incorporates the notion FEMALE, it also brings new information not predictable from *aunt* alone.

6 Meaning Extensions and Change

6.1 Established readings and nonce readings

Some of the alternative senses of a word are permanent and established features of the language, and we would expect them to be recorded in any dictionary worthy of the name. We may also assume that they are laid down in neural form in the mental lexicons of competent speakers of the language. This is the case with the two readings of *bank* discussed earlier. In a sentence like *Mary works in a bank*, we can say that the context "selects" a reading from among those that are permanently laid down, in the sense that only one of them yields a normal combination.

But take another case. Imagine a reception with a large number of guests who will later proceed to a dinner. There is no dining room large enough to accommodate all the guests, so they are divided into two groups. On arrival, each guest is presented with either a rose or a carnation. When it is time for dinner, the head waiter announces: *Will all roses proceed to Dining Room A, and carnations to Room B, please*. These uses of *rose* and *carnation* are perfectly comprehensible in context, but they are not an established part of the language, nor would we expect to find them in any dictionary, however complete. These are said to be nonce readings. How do they arise? Not by selection, but by coercion: if none of the established readings fits the context, then some process of sense-generation is triggered off, which produces a new reading.

A third possible effect of context is to enrich an existing reading, without producing a new sense. This is what happens to *teacher* in *John's maths teacher is on maternity leave*.

6.2 Literal and nonliteral readings

A distinction is often made between literal and nonliteral meanings of (polysemous) words, the assumption being that only one of the readings is literal. While at first sight this distinction seems intuitively clear, on closer examination it is not so straightforward. One thing is clear; a literal meaning must at least be an established one; the criteria for privileging one out of the set of established readings, however, are less clear. Dictionaries often order their entries in terms of chronological order of earliest attestation in the language. However, the earliest recorded meaning of a word does not necessarily strike speakers' intuitions as the literal meaning. There is no doubt, for instance, that the "die" meaning of *expire* predates the "cease to be valid" meaning, but (to my surprise) current British undergraduates, when asked to pick out the reading they intuitively feel to be the literal one, are virtually unanimous in selecting the "cease to be valid" reading. Another possible criterion is frequency in a language: one might reasonably expect the literal meaning to be the most frequent. Once again, however, this does not always accord with (strong) native speaker intuitions. For instance, few would dispute that the "have a visual experience" meaning of *see* is the literal meaning and "understand" an extended meaning, yet evidence from one very large corpus of English usage indicates that the latter is the more frequent. Two possibly more valid criteria are, first, the default meaning, that is, the one which comes first to mind when one is confronted by the word out of context, and, second, the reading from which the others can most plausibly be derived by standard semantic processes. The latter criterion is most reliable when there are more than two readings. Take the case of *position*, and three readings:

(i) "location in space"
(ii) "job in a large firm, etc."
(iii) "opinion"

Starting from (i), it is easy to see how (ii) and (iii) may have arisen by means of a process of metaphorical extension; but starting from either (ii) or (iii), there is no obvious way of deriving the other two.

6.3 Metaphor

There are a number of strategies for deriving one reading from another. Three will be illustrated here. The first of these is metaphor. Metaphor is essentially the projection of the conceptual structure appropriate to a familiar field onto a different and less familiar field, and depends for its effectiveness on a sufficient resemblance between the two fields for the projection to be intelligible. For instance, in the case of *expire*, projecting the notion of dying onto the life cycle of, say, a credit card, allows an immediately intelligible parallel to be drawn between death and the end of the period of usability of the card.

6.4 Metonymy

The second strategy of meaning extension is metonymy, which is based not on resemblances or analogies between items in different conceptual domains, but associations within a single conceptual domain. Referring to people wearing roses as *roses* is intelligible, not because of any structural parallels between the concept of a rose and the people designated, but because of the close association between the latter and roses in a particular situation. This usage is unlikely to become established. An example of metonymy that is so well established that we are hardly aware that it is nonliteral is *The kettle's boiling*.

6.5 Specialization and generalization

A third process which produces new meanings from old ones is a change of inclusiveness, widening or narrowing down the meaning. The meaning "take alcoholic beverage" of *drink* is derived from the meaning "imbibe liquid" by specialization; *meat* has become specialized from meaning food of any kind to "animal flesh used as food"; *handsome* in Jane Austen seems to apply indifferently to men and women, but became specialized later to apply primarily to men, acquiring a particular nuance in its application to women; *interfere* in Jane Austen seems rather like current *intervene*, acquiring its disapproving connotations at a later date. The meaning of *cat* which includes lions, tigers, ocelots, and jaguars is derived by generalization from the meaning which includes only *felis domesticus*.

The loss of a whole sense may be considered a type of specialization: an example of this, again from Jane Austen, is *direction*, which has lost the reading "address" (as on a letter).

6.6 Amelioration and pejoration

Amelioration is when a neutral or pejorative term becomes commendatory, and *pejoration* is when the reverse movement occurs. The latter seems to be by far the more frequent, and is particularly prone to happen to words referring to women. Examples of words undergoing pejoration are: *madam*, *mistress*, *courtesan*, *wench*, *tart*, and so on. (The only two words referring to women that Jane Miller in *Womanwords* signals as having ameliorated are *jilt* which originally meant a prostitute, and *bat*, which "has lost its negative sexual connotations)."

7 Larger Groupings of Words

7.1 Word fields

We have already seen that the vocabulary of a language is not just a collection of words scattered randomly through semantic space: it is at least partially structured by recurrent sense relations.

In some areas of the vocabulary the sense relations unite groups of words into larger structures, known as *lexical fields* (or *word fields*). We shall look briefly at examples of one type of larger structure, namely, *lexical hierarchies*. Hierarchies may be *nonbranching*, as in Figure 12.3, or *branching*, as in Figure 12.4.

If these are lexical hierarchies, then A, B, C, etc. represent word senses, and the lines joining them represent sense relations. The following are examples of nonbranching hierarchies (turning them on their side for convenience):

(i) general – colonel – major – captain – lieutenant – etc.

(ii) ocean – sea – lake – pond – puddle

(iii) scorching – hot – warm – lukewarm – cool – cold – etc.

(iv) tertiary (education) – secondary – primary – preschool

A structural necessity for a branching hierarchy is that the branches must never converge. There are two main sorts of branching lexical hierarchy, which are called *taxonomies*, which are structured by hyponymy and incompatibility and *meronomies*, which are structured by meronymy and co-meronymy. Figure 12.5 illustrates a fragment of a taxonomy (valid for British English).

It will be appreciated that the hierarchy illustrated in Figure 12.5 can be extended upward, downward, and sideways, ultimately producing a huge structure encompassing all living things. Most taxonomies are much more fragmentary than this, covering such areas as vehicles, buildings, clothes, and so on.

Figure 12.3

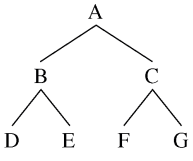


Figure 12.4

A taxonomy typically has well-defined levels: in Figure 12.4, A is at Level 1, B and C are at Level 2, D, E, F, and G are at Level 3. One level of a taxonomy has a special status and is known as the *basic level*, as it seems that the structure is organized around this level. It is the level at which the "best" categories occur. This means that resemblance between fellow members and

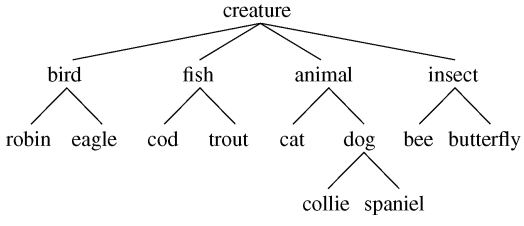


Figure 12.5

Note: A word of explanation is perhaps needed regarding the position of the word "animal" in the above hierarchy, which for some non-British speakers of English, who feel that "animal" and "creature" are synonyms, and that "animal" subsumes "bird," "fish," "insect," and so on, may seem anomalous. This is not so in British usage: the *Collins Handguide to the Wild Animals of Britain and Europe* (Ovenden, Corbet, and Arnold 1979), a field guide to identification, includes mammals, amphibians (such as frogs and newts), and reptiles (such as snakes and lizards), but no birds, fish, or insects (which are covered by sister volumes). I am not aware of any other language which has a term with exactly this denotation.

distinctiveness from members of sister categories are both maximized. Take the categories ANIMAL, DOG, and SPANIEL. Animals are distinct enough from nonanimals, but their overall degree of mutual resemblance is relatively low; in the case of spaniels, they resemble one another highly, but differ in only minor ways from other types of dog; in the category DOG, on the other hand, mutual resemblance and distinctiveness are both high. Basic level items are also the default names for things, the names we use for simple, everyday reference. Suppose A hears a noise in the garden and asks B what it is. B looks out of the window and sees a spaniel. What does he say? Any of the following would be true:

(i) Oh, it's just an animal.
(ii) Oh, it's just a dog.
(iii) Oh, it's just a spaniel.

In the absence of special circumstances, (ii), which contains the basic level item, is by far the most likely.

Some examples (in capitals) of basic level items are as follows:

a. vehicle – CAR – hatchback
b. fruit – APPLE – Granny Smith
c. living thing – creature – animal – CAT – Manx cat (British)
d. object – implement – SPOON – teaspoon

The other main type of branching lexical hierarchy is the part–whole variety. (In Figure 12.6, as in Figure 12.5, only some of the branches are shown.) Part–whole hierarchies are just as numerous in the vocabularies of natural languages as are taxonomies. They differ from taxonomies in a number of respects, but perhaps the most significant difference concerns structural levels: meronomies tend to have no, or only weakly developed, levels, hence there is no equivalent to the basic level of a taxonomy.

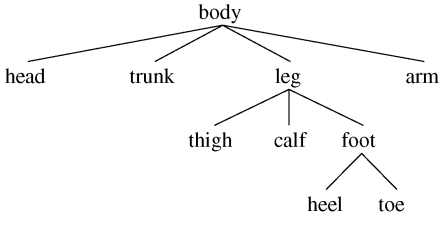


Figure 12.6

7.2 Word families

Another type of grouping of associated words is the *word family*. Most complex lexemes are built up out of a *root* and one or more *derivational affixes*. For instance, the word *undress* is composed of the root *dress* and the prefix *un-*; the noun *re-entry* is composed of the root *ent(e)r*, the prefix *re-* and the suffix *-y*. A word family is composed of all the words derived from a given root. For instance, the following all belong to one word family:

nation	national (adj.)	national (n.)
nationally	nationalize	denationalize
nationality	nationalism	nationalist
international	transnational	nationhood
(etc.)		

As native speakers, we have a quite complex knowledge of which derivations are possible, and what they mean. We know, for instance, that although a *painter* is someone who paints, and a *painting* is the concrete end result of a painter's efforts, a *screwdriver* is not someone who inserts screws, but an appliance for doing this, and a *killing* is not the concrete end result of the process denoted by *kill* (i.e., a corpse), but an instance of the act itself. We also know that whereas a *diner* may be someone who is dining, it can also be a place where one dines, and we know that there is no parallel reading for *painter*.

7.3 Domain-specific vocabulary

Another type of word grouping controlled by native speakers is the vocabulary appropriate to a particular situation, for instance, a race meeting:

1	2	3	4
horse	jockey	course	bet
favorite	owner	race	run
form	trainer	fence	win
odds	bookmaker	winner's enclosure	jump
colors	steward	starting-gate	take a fence
winner	stable-boy	heavy going	come up on the inside
handicap		reins	fall
yearling			disqualify
			scratch
			etc.

A broad grouping like this is composed of a number of nested subdomains, such as weighing in, saddling, starting, running the race, laying bets, and so on.

7.4 Layers of vocabulary

We shall use the expression *layers of vocabulary* to refer to much larger groupings of words, each of which will incorporate many structures like those described, which are confined to certain areas of usage. For instance, there are technical vocabularies, such as those used by art historians, or doctors, when communicating with others of their kind. There are also collections of words which associate together at different levels of formality. These are effortlessly called up in appropriate contexts, and they must be somehow linked in storage.

7.5 The mental lexicon

Each of us has in our cognitive system some kind of inventory of all the words that we know, together with all the information – semantic, grammatical, and phonetic / graphic – necessary for their correct use. Estimates of the number of words known by an average adult speaker vary from 150,000 to 250,000 (see, for instance, the discussion in Aitchison (1987: 5–7)). This represents a vast quantity of information.

The inventory is accessed via written or spoken forms every time we hear or read something in a language we know, and via some kind of semantic representation every time we produce language (recall that, because of widespread synonymy and polysemy, the mapping between forms and meanings is not one-to-one, but many-to-many). Although much is known, the details of representation and processes of use are still very imperfectly understood; nonetheless, the astonishing speed at which words are retrieved and identified – within about a fifth of a second from the start of the word, for spoken language – points to a highly efficient and organized storage system.

Every person's mental lexicon is different from everyone else's, yet by and large we manage to understand each other; this presumably indicates an adequate degree of overlap between individual lexicons.

7.6 Vocabularies

In addition to the mental lexicons of individual speakers of a language, it is possible to think of the total lexical stock of a language, which covers all its speakers, including those belonging to distinct speech communities, and including those who are now dead. Of course, the boundaries of such an entity are very vague, and will differ according to the purposes of the compilers and users (how far back in time do we go? how many dialect, or specialized technical forms do we include?, etc.). The natural home for such a vocabulary is the dictionary, and the natural way of drawing it up is by studying corpuses. The contents of a dictionary do not correspond to the contents of the mental lexicon of any single speaker, nor do they represent what is common to all speakers. However, every entry must be justified by some degree of common ownership in one or other of the subcommunities using the language.

8 Conclusion

We have now surveyed, at least in broad outline, the whole domain of words in language, from the detailed properties of individual words, through relational properties along the major paradigmatic and syntagmatic axes, to communities of words, large and small, tightly or loosely structured.

It should be borne in mind, however, that many detailed aspects of the lexicon are still only imperfectly understood, and of these, a number are currently the object of intense research activity. Two recent major stimuli to research on the lexicon are worth mentioning. The first has been the advent of powerful computers, and the attempt by computational linguists to develop programs capable of "understanding" natural language texts. The syntactic problems have proved relatively tractable; the big problem has turned out to be the lexicon – deciding what a computer must "know" about word meanings and how they are to be represented. The second major stimulus has been the development of large-scale corpuses of spoken and written language (together with tools for processing them), which allow an accurate picture to be gained of how words are actually used. This has, among other things, revolutionized lexicography, and no doubt the full effects of both stimuli have yet to be seen.

REFERENCES

Aitchison, J. (1987). *Words in the Mind*. Oxford: Basil Blackwell.

Cruse, D. A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.

Lyons, J. (1963). *Structural Semantics*. Cambridge: Cambridge University Press.

Mackin, R. (1978). On collocations: words shall be known by the company they keep. In P. Strevens (ed.), *In Honour of A. S. Hornby* (pp. 149–65). Oxford: Oxford University Press.

Ovenden, D., Corbet, G., and Arnold, N. (1979). *Collins Handguide to the Wild Animals of Britain and Europe*. London: Collins.

13 Syntax

MARK C. BAKER

1 The Domain of Syntax

Syntax is the branch of linguistics that studies how the words of a language can be combined to make larger units, such as phrases, clauses, and sentences.

As such, syntax is deeply concerned with the relationship between the finite and the infinite. Languages have a finite number of basic words, but those words can be put together to make an infinite number of sentences. Thus, one can buy a more or less complete dictionary of English, but it is unthinkable to buy a reference book that would list all the sentences of English. This is because there are rules and patterns that can be used dynamically to create and understand new English sentences. Syntax is the study of these rules and patterns. For example, readers of this article are probably aware of having encountered the English words *this*, *loves*, *elephant*, and *eggplant*, and these are listed in their dictionaries. However, they are probably not aware of having encountered the English sentence in (1) before, and they are unable to "look it up":

(1) This elephant loves eggplant.

Nevertheless, English speakers have no hesitancies about recognizing this as English, or about being able to understand it and use it appropriately. Moreover, they feel very differently about (2), even though it is made up of the same elements:

(2) *Loves elephant eggplant this.

In this arrangement, the combination is not well formed, but rather is *ungrammatical* (indicated by the * symbol).

Syntax is primarily concerned with whether a sentence is "properly put together" rather than whether it is meaningful, or silly, or bizarre. Thus, it is concerned with the difference between (1) and (2), more than in the difference between (1) and the example in (3b) (Chomsky 1957: Chapter 2).

(3) a. Those hippopotamuses hate asparagus.
b. #This eggplant loves elephants.

These sentences are all instances of the same pattern from a syntactic point of view, even though they mean different things – (3b) being nonsensical, and unacceptable for that reason. What new sentences formed by the patterns of syntax actually mean is the primary concern of semantics.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

However, aspects of how words are put together do influence how the sentence is interpreted; hence these issues can come into the domain of syntax as well. There is thus a complex interaction between the disciplines of syntax and semantics, and it is controversial whether the topics can truly be distinguished.

Two other branches of linguistics that are closely related to syntax are morphology and discourse analysis: morphology because it builds the words that are the starting point of syntax, and discourse analysis because it involves the combination of sentences into even larger entities, such as texts. Again, it is somewhat controversial to what extent syntax is a distinct topic from these others. It is conceivable that sentences are built in accordance with the same patterns and procedures as words or texts are. If so, syntax is not a separate topic from morphology or discourse analysis. Be that as it may, there can be no doubt that some kind of syntax exists, and that it involves putting the finite building blocks (morphemes or words) made available by a given language together into an infinite number of representations, which can express an infinite number of thoughts.

2 The Chomskyan Perspective

The study of syntax has flowered in the modern period (since 1950), having been given new impetus by the work of Noam Chomsky. In this section, I begin the discussion by considering (part of) his founding influence on the field.

2.1 The use of formal mathematical tools

One small but essential part of Chomsky's contribution has been to bring certain insights from the study of formal languages in mathematics to bear on the fundamental problems of syntax. For example, the mathematical concept of a recursive function sheds crucial new light on how a finite number of words can be pieced together to make an infinite number of sentences (Chomsky 1975: Chapters 7 and 8). A recursive function is one that is defined in terms of itself – a "circular" definition that succeeds in saying something substantive. For example, suppose that English contains the following rules:

(4) a. A sentence (S) consists of a Noun Phrase (NP) followed by a Verb Phrase (VP).
(S → NP+VP)
b. A VP consists of a verb (V), possibly followed by an NP and/or a clause (CP).
(VP → V, or V+NP, or V+CP, or V+NP+CP)
c. A CP consists of an S, possibly preceded by a complementizer (C)
(CP → S or C+S)

English also includes the following words:

(5) Noun phrases: *John, Mary, Bill*
Verbs: *believes, likes*
Complementizer: *that*

This is a very small vocabulary, and a short list of rules/patterns. However, the set of rules is *recursive* in that a sentence (S) contains a VP, and one of the things a VP can contain is a clause (CP), which always contains a new S. This S in turn contains its own VP, and that can contain yet another S. And so on. Thus, with only these rules one can already make an infinite number of grammatical English sentences, depending on how many times one puts a new S inside the verb phrase:

(6) Mary likes Bill.
John thinks that Mary likes Bill.
Bill thinks that John thinks that Mary likes Bill.
Mary thinks that Bill thinks that John thinks that Mary likes Bill.
John thinks that Mary thinks that Bill thinks that John thinks that Mary likes Bill…

Each of these sentences expresses a distinct idea – one that is potentially useful in the world of high school courtship. There is clearly much more to English than this, but it shows how the mathematical idea of recursion can elegantly address the apparent paradox that people who have a finite amount of knowledge of English stored in their minds can produce and understand an infinite number of English sentences.

2.2 *The goal of accuracy and explicitness*

A second foundational contribution of Chomsky and the Cognitive Science revolution of the 1950s and 60s has more to do with the spirit of syntactic inquiry than its theoretical machinery. This is the ethos that the syntactic rules and patterns of a particular language should be stated explicitly and precisely, with (ideally) nothing being taken for granted or left to the imagination (Chomsky 1957: Preface). To see why this is important, consider the difference between telling a friend to do something and programming a computer to do something. Interacting with your friend is generally easier, because you expect the friend to bring a great deal of common sense to bear on the request, filling in the details and interpreting what you say in the light of circumstances. In contrast, computers have no inherent common sense; they do all and only what they are told in the most literal fashion. Although this can make working with computers exasperating, it can also be educational, because programming a computer to do a task forces one to take stock of exactly what goes into that task, with no question-begging or step-skipping. This turns out to be highly relevant to the study of language. The syntax of a language can seem deceptively simple, particularly in a monolingual situation. For example, if we all speak English, it is very easy for us to fill in details without realizing it. If, however, one takes up the challenge of describing the patterns and principles of syntax to the point where one could in principle program a computer to judge sentences, manipulate them, and interpret them the way we do, one is forced to seriously investigate many issues that are otherwise easily overlooked. Sometimes this may seem obsessive to an outside observer: syntacticians can seem to be giving needlessly complex analyses when it is obvious that we say (1) and not (2) because "(1) sounds better" or "(1) makes more sense." But in fact syntacticians are at least as interested in exactly what goes into linguistic "common sense" as they are in the salient and arbitrary rules of grammar that get most of people's attention. Many important discoveries have been made by taking this perspective; some examples are illustrated below.

2.3 *The goal of simplicity and generality*

A more controversial feature of the Chomskyan legacy in syntax is its emphasis on simplicity, elegance, and generality in linguistic analysis. The syntactician is continually challenged to filter out the complexities in a linguistic description and look for what is most simple and general underneath the morass of detail. This has been a characteristic of Chomskyan research for decades, but it has been given increased prominence in the last 20 years within the so-called "Minimalist Program" (Chomsky 1995). In part, this drive for elegance and generality is just good scientific practice, which seeks to maximize theoretical parsimony in every domain. But for some Chomskyans, it can be more than this: an intuition that there is something special

about human language within the mess and complexity of the biological world. For example, the drive toward simplicity can be motivated by a desire to have a theory of human language that is compatible with its having appeared recently and suddenly in evolutionary time. Therefore, it is thought, our biological capacity for language cannot have accumulated too much complex construction-specific gadgetry.

To see how these considerations can come into play, consider the phrase-building rules in (4). Early treatments of English syntax collected more and more of these rules, and the rules became longer and more detailed so as to account for more data, as is natural. But the details of these rules threatened to obscure more general patterns. For example, rule (4b) states that every verb phrase begins with a verb, among other things. Other patterns of English imply that prepositional phrases (PP) begin with prepositions, noun phrases begin with nouns, and adjective phrases (AP) begins with adjectives, as illustrated in (7).

(7) a. in Rome PP → P NP
b. pictures of Rome NP → N PP
c. proud of Rome AP → A PP

A clear generalization can be drawn: every phrase in English starts with a word that defines the basic syntactic nature of the phrase (this is called the *head* of the phrase). The quest for simplicity and generality requires that this "coincidence" be given its due in the theory. The first step in this regard (Chomsky 1970; Jackendoff 1977) was to say that there is one very general phrase structure rule in English, of the form "XP is made up of an X followed by other phrases," where "X" is a variable that ranges across the various syntactic categories.

More recently, Chomsky has proposed doing away with even this schematic phrase structure rule, replacing it with two very basic principles, Merge and the Inclusiveness Condition (Chomsky 1995). Merge simply says that two syntactic objects can be joined together to make a larger syntactic object; for example the adjective *proud* and the PP *of Rome* can be merged to make a linguistic unit. The Inclusiveness Condition says that no properties can be added to a syntactic representation out of the blue; they must all be grounded in the lexical properties of the words that make up that representation. The syntactic character of the newly formed phrase *proud of Rome* must thus be inherited from the character of one of its parts; in this case, it is an adjectival phrase, because it contains *proud*, an adjective. In this way, some key features of the phrase structure rules in (7) can be deduced from very simple and general principles. (One feature of (7) that this minimalist reduction does not capture is that the head is the first word in its phrase in English; we return to this below.)

There is a tension between the goals of syntactic research. The desire for explicitness and accuracy tends to lead the linguist to put more and more detail and complexity into a linguistic theory, whereas the desire for simplicity and generality can lead one to downplay such detail. For example, some insight is gained by collapsing the explicit phrase structure rules of English into a single schema like XP → X…, but some accuracy is lost as well. For example, not all heads appear with the same range of phrases. Verbs can be followed by NPs, PPs, or CPs, as we have seen, whereas nouns can be followed by PPs or CPs but not NPs (*claims about Chris*, *claims that Chris died*, but not **claims Chris*). These details could be encoded in individual phrase structure rules, but they are not captured in one general phrase structure schema, nor by its minimalist reduction. Syntactic theory thus needs to seek other explanations for these differences. This is a healthy pressure precisely to the degree that new and better explanations of the details are forthcoming, deducing them from some generality along another dimension of analysis. Linguists currently disagree as to how much generalization across syntactic patterns is actually warranted, and this is an area of ongoing research (see Section 5).

The broadly Chomskyan approach to syntax is often called *formal* or *generative*: "formal" because it uses mathematics-style formalisms and definitions in presenting its analysis, such as the

recursive rules in (4); "generative" because it seeks to explicate how fluent speakers can generate new expressions of a language. Many syntacticians accept these labels for their work, while others have reacted against what they see as some deficiencies and excesses of this approach. However, it seems fair to say that even those that react most loudly against the Chomskyan approach have been positively influenced by its most basic ethos and by some of what has been learned from that perspective. For the purposes of this article I attempt to foreground these areas of relative agreement.

3 Lessons of Syntactic Research

What are some of the substantive discoveries that have been made by adopting this approach to syntax? Beyond the details of particular results, one can identify at least three broad themes: that syntax is a vast topic, that constraints are central to syntax, and that there is a large component of syntax that is common to all human languages. I discuss the first two themes in this section, and the third in Section 4.

3.1 The vastness of syntax

The first important lesson of several decades of research is that there is much more to syntax than anyone imagined. Natural languages turn out to be enormously complicated, once one adopts the "take nothing for granted" attitude of the computer programmer. We naturally think of an activity like playing chess as an epitome of intellectual achievement, mastered only by the brightest, whereas forming grammatical English sentences seems like a very ordinary activity, done almost effortlessly by practically everyone. However, the experience of computer science shows vividly that the real complexity of the tasks is the reverse. Steady progress has been made in programming computers to play chess, so that they now beat even the best human players. In contrast, there is still no computer system that can match the average five-year-old's ability to judge, generate, and interpret English sentences, despite enormous resources having been devoted to the task. Even the longest grammars of the best-studied languages are not complete descriptions, and new discoveries about languages like English and French are made on a regular basis.

To see where some of the complexities come from, let us return to the notion of phrase structure. Rules like S → NP+VP and VP → V or V+NP express in a succinct way several facts (Chomsky 1965: Chapter 2). First, they express the fact that all normal sentences must have at least a subject NP and a verb (except for imperatives, and certain elliptical expressions). If the subject is omitted, the sentence is deviant:

(8) a. Mary likes the dress.
    b. *Likes the dress.

Second, these rules express the fact that the subject appears before the verb, and the object appears immediately after it. Thus, *The dress likes Mary* is a silly sentence with a completely different meaning from (8a); it cannot be understood as having the object before the verb and the subject after it, but only as *the dress* being the subject and *Mary* being the object. Other logically possible rearrangements of the words are simply ungrammatical (**Mary the dress likes*, **Likes Mary the dress*, etc.). Third, the phrase structure rules express the more subtle fact that the object and the verb together form a tighter unit (the verb phrase) in English, with the subject attached to the sentence more loosely. As a result, in some special contexts one can put the verb and the object together at the front of the sentence, leaving the subject behind (see (9a)). In contrast, one cannot put the subject and the verb at the front, leaving the object behind (9b).

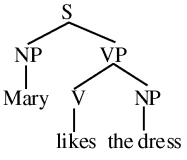
(9) I told you that John would like that dress, and indeed…

a. …Like that dress, John does.

b. *…John like, does that dress.

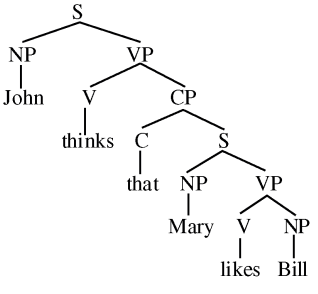
Putting these pieces together, we can diagram the structure of a simple English sentence as follows:

(10)



Similarly, the second sentence in (6) would have the phrase structure diagram in (11); note the recursion, with one S embedded inside another.

(11)



Already some complications arise. The given phrase structure rules account for the ungrammaticality of (8b), where the subject has been omitted. Now it is just as bad to leave out the object of (8a):

(12) *John likes.

However, unlike (8b), (12) can be generated by the phrase structure rules in (4). The difference is intentional. It is motivated by the fact that (12) becomes grammatical when another verb is substituted for *likes*; for example, one can say *John smiles*. In contrast, sentences like (8b) are impossible whatever verb is chosen (**Arrived the dress, *Tore the dress*, etc.). It is a property of the particular English verb *like* that it needs an object, whereas it is a property of the English language as a whole that sentences need to have subjects.

Following the minimalist urge, we can factor these properties out of the specific phrase structure rules, and state them as general conditions instead. The fact that all clauses must have a subject is often called the *Extended Projection Principle* (Chomsky 1981). The fact that an object must be present if and only if a particular verb calls for one semantically is often called the Theta Criterion (Chomsky 1981).

There is also much more to syntax than phrase structure rules and the principles that underlie them. There are also systematic relationships among sentence types that need to be captured, for which Chomsky originally proposed the formal device of the *transformational rule* (Chomsky 1957: Chapter 7). These rules change one phrase structure representation into another. For example, consider the following English sentences:

(13) a. Mary will like the dress.

b. *Mary will like what? (bad with neutral stress on *what*)

c. What will Mary like?

d. *What will Mary like the dress?
e. *The dress will Mary like.

English has a special set of words that can take the place of a missing piece of information in order to form a question; these include *who*, *what*, *where*, *which*, and others. However, (13b) shows that it is not enough to simply put one of these words in place of the desired piece of information. Rather, the question word must appear at the beginning of the sentence, as in (13c). Moreover, (13c) is a striking exception to the generalization that the verb *like* must be immediately followed by an object. In this environment, *like* must *not* have an object after it (see (13d)). One natural way to describe this network of facts is to say that the initial structure of (13c) is like that of (13a), with *what* substituted for *the dress*. However, the structure is "transformed" by a rule that can be stated something like this:

(14) Move question words to before the first word of an S that contains them.

This rule applies to question words, but not to other NPs; hence (13e) is not possible, even though it is parallel to (13c).

Transformational relationships like these are relatively common. Indeed, (13) illustrates a second one. Notice that the placement of the future auxiliary *will* is different in declarative sentences like (13a) and interrogative ones like (13c). In declarative sentences, it comes after the subject and before the main verb, whereas in interrogative sentences it comes between the question word and the subject. This change in word order is required in main clauses, and one cannot have a second auxiliary of the same type in the usual position between the subject and the verb:

(15) a. *What Mary will like?
b. *What will Mary will/may like?

Again, a movement transformation is a simple way to account for these facts:

(16) Shift the auxiliary verb to the left of the subject NP in main clauses interpreted as questions.

There may also be processes that apply to phrase structures that do not fall under the general category of movement. Suppose that we expand our miniature English dictionary to include NPs that are pronouns, such as *he*, *she*, and *it*. Then the phrase structure rules generate examples like (17):

(17) John thinks that he likes Mary.

This sentence is ambiguous. Depending on the context, it can easily be interpreted as meaning that John thinks that John himself likes Mary, or that John thinks that some other male under discussion (say Bill) likes Mary. This an essential feature of how competent English speakers understand and use sentences with pronouns. Suppose, then, that each distinct person referred to is associated with a unique number, written as a subscript to the NP(s) that refer to it. The ambiguity in (17) can now be captured by a rule like the following (Chomsky 1973).

(18) Copy the referential index of an NP onto a pronoun as its referential index.

This is an interpretive rule, rather than a transformational rule. It is optional: one meaning of (17) comes from applying it, and the other meaning comes from not applying it.

(19) a. John_3 thinks that he_3 likes Mary_4, *or* (by (18))
b. John_3 thinks that he_2 likes Mary_4.

Just as the transformational rules in (14) and (16) apply only to a specific set of items (*wh*-words, auxiliary verbs), so (18) applies only to pronouns. Thus, if the man named John also happens to

be a sentimental fool, one can use either the NP *John* or *the sentimental fool* to refer to him. However, if one replaces the pronoun *he* in (17) with the NP *the sentimental fool*, the sentence can only have a meaning like (19b), where three distinct people are being discussed. This is because *the sentimental fool* is not a pronoun, and therefore rule (18) does not apply to it.

The existence of transformational and interpretive rules is more controversial than the existence of phrase structure. These facets of a linguistic analysis can be eliminated if one makes up new, more complex phrase structure rules that will generate the desired sentences directly (Gazdar et al. 1985). Some syntacticians prefer this alternative, often for reasons of computational tractability or formal elegance. However, there is no doubt that there need to be *some* general mechanisms for expressing relationships among the different parts of a syntactic structure – such as the relationship between the sentence-initial question word in (13c) and the fact that the sentence has no NP in the object position of the VP.

3.2 The centrality of constraints

All this is far from exhausting the domain of syntax, however. Indeed, it merely sets the stage for perhaps the most interesting and significant discovery of all: the existence of *constraints*.

When the rules of syntax are stated in their simplest form, they typically "overgenerate," producing certain ungrammatical sentences along with many grammatical ones. Therefore, syntacticians have proposed a system of constraints that prevent these rules from running wild.

As an example, let us return to rule (14), which moves question words to the front of sentences. This is a very general phenomenon, but it turns out to be far from exceptionless. Consider the following more complex examples:

(20) a. John will think that Mary likes the dress.
b. John will think Mary likes the dress.
c. What will John think Mary likes –?
d. What will John think that Mary likes –?
e. Who will John think – likes the dress?
f. *Who will John think that – likes the dress?

Examples (20a) and (20b) are simple sentences, formed by the phrase structure rules in (4). They differ only in whether the complementizer *that* is present or not: in (20a), the CP inside the VP headed by *think* is expanded as C+S (where *that* is C); in (20b), this CP is expanded simply as S. Examples (20c) and (20d) are the result of generating *what* as the object of the embedded S and moving it to the front; both are grammatical, as expected. Examples (20e) and (20f) are the result of substituting *who* for the subject NP of the embedded sentence and moving it to the front. Now we find something unanticipated: (20f) is ungrammatical for most speakers, even though our rules can construct this sentence just as easily as the others. There seems to be an additional constraint at work here, which can be stated as follows (Chomsky and Lasnik 1977):

(21) *Complementizer Constraint*
A sentence is ungrammatical if a complementizer comes immediately before a verb (in languages where overt subjects are generally obligatory).

Since only (20f) has the forbidden sequence of words … *that likes* …, it alone is ruled out by (21).

There are other constraints on question movement as well. Either a subject or an object can be replaced by a question word, which then moves to the front of the clause. However, a difference between subjects and objects appears when one tries to question a subpart of the NP. When a question word replaces part of an object NP and moves to the front, the result is often acceptable, as in (22b).

(22) a. You saw [a picture of John].
b. Who did you see [a picture of –]?

However, when a question word replaces a similar part of a subject noun phrase, the result is unacceptable, as in (23b).

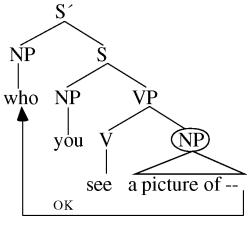
(23) a. [A picture of John] disturbed you.
b. *Who did [a picture of –] disturb you?

To account for the badness of (23b), syntacticians have stated the following condition (Huang 1982):

(24) *The Condition On Extraction Domains* (CED):
A phrase X can move out of a phrase Y only if Y is immediately contained in a Verb Phrase.

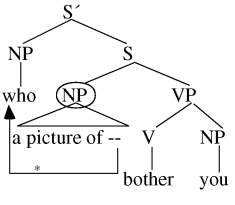
Since the object is inside the VP, as shown by the phrase structure diagram in (25), the movement in (22b) is possible:

(25)



However, the subject NP is not inside the Verb Phrase; therefore question words cannot move out of the subject:

(26)



The CED thus correctly differentiates between good sentences like (22b) and bad ones like (23b).

There are also interesting constraints on the pronoun interpretation rule in (18). Suppose that the object of a sentence is a pronoun. That pronoun can easily be understood as referring to something inside the subject noun phrase:

(27) John's$_2$ mother$_3$ loves him$_2$.

The situation is not symmetrical, however. If we put a pronoun in the subject position, it cannot normally be understood as referring to something contained in the object. Thus, *he* cannot be John in a sentence like (28), but must be some third person.

(28) *He$_2$ loves John's$_2$ mother$_3$.

One might think that this is simply a matter of linear order: a pronoun can never refer to a noun phrase that comes after it. However, this is not the whole truth. For example, in a sentence like (29), the pronoun *his* comes before *John*; nevertheless, *his* can be understood as referring to John,

for most English speakers. (This interpretation is easier to get if one says the sentence with main stress on the verb *loves*, rather than on *John*.)

(29) (?)$\underline{\text{His}}_2$ mother_3 loves $\underline{\text{John}}_2$.

We thus need to add a condition to (18), which by itself would imply that pronouns can always refer to any other NP in the same sentence (Reinhart 1983; Lasnik 1989). It helps to first define the term *c-command*:

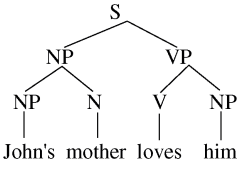
(30) C-command:
An element X *c-commands* another element Y if the first phrase which properly contains X also properly contains Y.

Given this, we can state the condition on pronouns as follows:

(31) *Disjoint Reference Condition* (DRC):
A pronoun X may not refer to the same thing as (have the same index as) a nonpronominal NP Y if X c-commands Y.

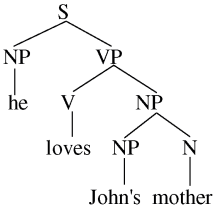
This explains the facts as follows. Consider (32), which is the structure of (27).

(32)



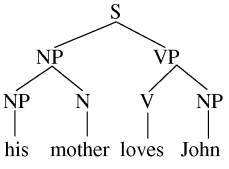
First we must check if the pronoun *him* c-commands the NP *John*. The first category that properly contains *him* is the VP, and *John* is outside this VP. Thus, the pronoun does not c-command *John* in this structure. Therefore, the DRC does not apply, and (18) can; the pronoun may refer to the same person as *John*. In contrast, the structure of (28) is (33).

(33)



Here the first category that contains the pronoun *he* is S, the whole sentence, and S obviously contains the NP *John* as well. Therefore, the pronoun does c-command the noun phrase in this structure, so the pronoun cannot refer to the same thing as *John*, by the DRC. Finally, (34) is the structure of (29).

(34)



Here the pronoun *his* is not the whole subject, but only a part of the subject. Thus, the first phrase which properly contains *his* is the subject NP. This subject NP does not include the object *John*. Therefore, the pronoun does not c-command the NP in this example, and coreference is possible.

Therefore, we see that a condition defined over the phrase structure of the sentence accounts for the usage of pronouns in English better than a simplistic rule that says that a pronoun must come after the NP it refers to.

Much of the importance of these constraints comes from the fact that they regulate the operation of more than one transformational relationship (Ross 1967; Chomsky 1973; Chomsky 1977). For example, in addition to the transformations mentioned above, English has a rule of relative clause formation that makes noun phrases out of sentence-like structures. A simple example is:

(35) Mary likes the dress → the dress that Mary likes –

This is not the same process as question formation, but it is somewhat similar in that an NP (here *the dress*) comes to be at the beginning of the clause, and there is a gap after the verb *likes*, where a noun phrase normally has to be. Thus, this too can be thought of as a transformational movement process. Now, given the way the constraints are set up, we automatically expect relative clause formation to have the same limitations as question formation. This is correct. Like question movement, relative clause formation cannot create a configuration where the complementizer *that* comes immediately before the verb, as shown in (36) (compare (20)).

(36) a. the dress that John thinks Mary likes –
b. the dress that John think that Mary likes –
c. the woman that John thinks – likes the dress
d. *the woman that John thinks that – likes the dress

Relative clause formation also obeys the CED, such that part of the object can be moved, but not part of the subject. Example (37a) is thus parallel to (22b) and (37b) is parallel to (23b).

(37) a. the man that you saw a picture of –
b. *the man that a picture of – disturbed you

This is an important finding, because it shows that the constraints should not be built into the individual transformations themselves; rather they have a semi-independent existence. In principle, one could write the Complementizer Constraint and the CED into the formulation of the question movement transformation. The result would be a rule that is rather complex and inelegant (not minimalist). Furthermore, the very same complexities would have to be written into the relative clause transformation – and into a number of other transformations as well. It is clearly better to factor these out as separate conditions. This has led to the view that the constraints are really more fundamental than the transformations themselves, a profound shift from seeing language as fundamentally a list of construction rules to seeing language as a system of constraints that must be satisfied (Chomsky 1981).

In summary, we have seen that syntax is surprisingly complex. It includes much more than a simple statement about basic word order, which is all that many traditional descriptive grammars say about the topic. Rather, it includes phrase structures, transformations defined over those phrase structures, and constraints that limit the transformations defined over the phrase structures in systematic ways. But (paradoxically) syntax is also simple, in that the complex structures we see are those that satisfy a collection of relatively simple constraints. These constraints are arguably the key to syntax. It has been conjectured within the Minimalist Program that syntax really reduces to "build anything" and "move anything," as long as the constraints are satisfied – which they are not, except in the rather specific conditions described by traditional descriptive rules. Indeed, the Extended Projection Principle and the Theta Criterion mentioned in Section 3.1 should join the list of constraints discussed here. These are general constraints that govern the operation of structure building (and hence potentially replace phrase structure

rules), whereas the Complementizer Constraint and the CED are general constraints that govern the operation of transforming structures (and hence potentially replace transformational rules).

4 The Similarities and Differences among Human Languages

The third major lesson of contemporary syntactic research is that human languages are strikingly similar in many respects. Indeed, there has been an accelerating interest in this aspect of syntactic theory in the twenty-first century, related in part to the effort to document and preserve endangered languages. I thus single this out for some extended discussion.

It is certainly not one's first impression that the syntaxes of all human languages are similar. Some 6,000–7,000 languages are currently spoken in the world, many of which had little contact with each other until recently. When speakers of an Indo-European language first start to learn a native American language, or an Australian aboriginal language, they are usually more struck by difference than by similarity. However, the differences turn out to be largely on the surface.

We have seen that when one tries to take nothing for granted one begins to discover many unsuspected intricacies of English grammar. Not surprisingly, when one begins to study other languages from the same perspective, one finds unsuspected intricacies in those languages too. What is surprising is that the intricacies turn out to be largely the same, even across languages that are historically unrelated. Moreover, this seems to be truer the further one goes into the linguistic analysis. Especially at the level of syntactic constraints, it is striking that what is originally discovered in languages like English and French often shows up in some recognizable form in other languages. The syntax of natural languages is thus not only vaster than we thought, but it is also more similar than we thought. In this section, I illustrate this theme with three case studies: Edo, Japanese, and Mohawk.

4.1 The Syntax of Edo

What is striking about the Edo language, spoken in Nigeria, is that it looks very much like English, even superficially (Stewart 2001). This is striking because the language had no contact with English until modern times.

First of all, the basic phrase structure rules that work in English work also in Edo. For example, the subject comes before the verb, and the object comes immediately after it; if there is a PP, it comes after that, with the P before the NP. As in English, a subject noun phrase is obligatory in all clauses, but whether an object is needed or not depends on which verb is chosen. The object and verb form a relatively tight unit (the VP) which excludes the subject, as in English. It is also possible to have a clause inside the VP in Edo, as in English; this comes after the verb, and begins with a complementizer. Many of these facts are illustrated all at once in the complex sentence in (38).

(38)

Òzó	má	tá	wéé	írẹ̀n	ghá	rhiè	éfótò	Úyì	yè	né!é	ékpétìn.
Ozo	not	say	that	he	will	put	photo	Uyi	in	the	box.

'Ozo did not say that he will put a photo of Uyi in the box.'

The cumulative effect of these similarities is that a word-by-word translation of an Edo sentence usually makes a decent translation of the sentence as a whole into English, even for quite complex sentences like (38).

Edo also has striking similarities to English when it comes to transformations and conditions that are defined over them. For example, question words must be fronted to the beginning of the clause in which they appear in Edo, just as in English:

(39) a. *Òzó ghá guọ̀ghọ́ dèmwìn?
Ozo will break what

b. Dèmwìn nè Òzó ghá guó!ghó?
what that Ozo will break
'What will Ozo break?'

In English, the fronting of a question word is restricted by the CED, stated in (24). This constraint also holds in Edo. Example (40a) shows that one can generate a question word as the object of a clausal object and then move it to the front of the sentence as a whole. In contrast, (40b) shows that one cannot generate a question word inside a clausal subject and move it to the front of the clause (clausal subjects like this are perfectly fine in nonquestions).

(40) a. Dèmwìn nè Úyì tá wé!é òzó dé –?
what that Uyi said that Ozo bought.
'What did Uyi say that Ozo bought –?'

b. *Dèmwìn nè wèé Òzó dé – yèé Úyì?
What that that Ozo bought please Uyi.
'What did that Ozo bought – please Uyi?'
(intended meaning: 'What did it please Uyi that Ozo bought?')

Note that the judgments are the same for the English translations, an indication that the same constraint is at work in both languages.

Question movement in English is also subject to the Complementizer Constraint, stated in (21), which shows up when one tries to move an interrogative subject out of a clausal object. The sequence of complementizer *wee* followed immediately by verb is also avoided in Edo, as shown in (41a). Interestingly, the "fix" is different in Edo than in English. English avoids the forbidden configuration by omitting the complementizer, whereas Edo avoids it by filling the space left behind by the question word with a pronoun, as shown in (41b) (see Koopman 1982 on Vata, a related language).

(41) a. *Dòmwàn nè Úyì tá wèé – dé ímọ́tò.
Who that Uyi said that buy car.
'Who did Uyi say that bought a car?'

b. Dòmwàn nè Úyì tá wèé ó dé ímọ́tò.
who that Uyi said that he buy car.
'Who did Uyi say bought a car?'

This difference stems from the fact that Edo, unlike English, requires embedded clauses to have a complementizer; in terms of explicit phrase structure rules, CP → C+S is the only possibility allowed in Edo (compare (4c)). This implies that (41a) cannot be fixed by leaving out *wèé*, and another solution must be found that is within the grammatical resources of the language. This shows that somewhat different grammatical patterns can be explained as accommodations to the same syntactic constraints, illustrating the point that the constraints on transformations are the most abstract level of linguistic analysis, but also the most universal.

It is easy to find more examples of both similarity and difference. For example, it is straightforward to show that Edo has pronouns like those of English, that these pronouns can undergo the coindexing rule in (18), and that the coindexing rule is limited by the Disjoint Reference Condition in (31) in Edo as in English. On the other hand, the examples in (39), (40), and (41)

show that Edo does not have the equivalent of English's Subject-Auxiliary Inversion transformation in (16): the future tense particle *ghá* does not move to between the question word and the subject the way *will* does in English. And yet it is notable that *something* must come between the question word and the subject even in Edo: in Edo it is an extra complementizer *nè* 'that'. Some linguists have thus proposed that there is a more abstract constraint that applies to both (all?) languages, to the effect that there is a syntactic position between the question word and the subject that must be filled by some function word (Grimshaw 1997). English responds to this constraint by shifting its auxiliary verb into that position, Edo by inserting a new word from the lexicon. Here again, the transformations are not the same across languages, but the constraints apparently are.

It is perhaps not surprising that *some* other language of the world is syntactically similar to English. What is very surprising is that so many of them are. With respect to word order and phrase structure in particular, it turns out that some 42 percent of languages of the world have essentially this word order (Dryer 2005).[1] That is a much bigger percentage than one would expect if phrase structure rules could vary randomly across languages. This degree of uniformity strongly suggests that the principles that underlie phrase structure in English are not specific to English, but apply to human languages more generally. The same is true for transformations and constraints on them. It is not known exactly how many of the 42 percent of languages with English-like word order also obey the CED and the DRC, but it is thought to be very high. Certainly it would be much bigger news within the syntax community if someone found a language in which these conditions did not hold than if they found another language in which they did.

4.2 The syntax of Japanese

In contrast, Japanese is interesting because it looks quite different from English and Edo. While a surprising 42 percent of languages have approximately the same word order as English, an almost-equal number (40 percent) have an alternative word order, and Japanese is one of these. Thus, the word-by-word translation of (42) in Japanese is all mixed up, nearly incomprehensible to an English (or Edo) speaker.

(42) Tara-ga Helen-ga Hannah-ni zibun-no syasin-o miseta to omotte iru.
Tara Helen Hannah-to self-of picture showed that thinking is
'Taro thinks that Helen showed a picture of herself to Hannah.'

But even when languages look the most different syntactically, closer analysis reveals important similarities. Japanese's word order is not randomly different from English's, but systematically different (Kuno 1973). Verbs come after their object NP or CP in Japanese, as shown in (42) and in the much simpler (43). Example (43) also shows that prepositions (postpositions) come after NPs when forming a PP in Japanese.

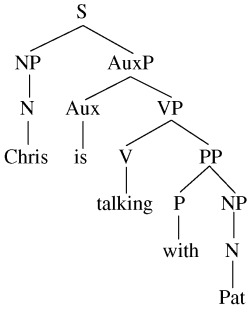
(43) John-ga Mary-ni hon-o yatta.
John Mary-to book gave
'John gave the book to Mary.'

We can also see in (42) that nouns come at the end of noun phrases in Japanese (*zibun-no syasin* 'picture of oneself', literally 'self-of picture'), and adjectives come at the end of adjective phrases (*Taro yori kasiko* 'smarter than Taro', literally 'Taro from smart').

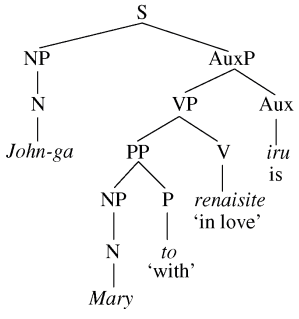
If we think of phrases in Japanese as being constructed by explicit phrase structure rules, in the style of (4), then every phrase structure rule in Japanese would be different from its English correspondent. But if we indulge the "minimalist" urge to simplify and generalize (see Section

2.3), then we recognize that the languages have much in common. Japanese has the same basic building blocks as English: it, too, distinguishes nouns, verbs, adjectives, and pre/postpositions. Japanese also groups its words together into phrases in the same way that English does; for example, the object NP and the verb form a VP in Japanese and hence end up next to each other, just as they do in English. Finally, the head of each phrase is in a consistent position in Japanese, just as in English. The one difference is that the head consistently comes at the end of the phrase in Japanese, whereas it comes at the beginning in English. Now recall that the minimalist reduction of phrase structure to Merge plus the Inclusiveness Condition implied that all phrases have heads, but it left open where the head would be. We now see that this is a welcome result, because this in fact varies across languages, while other aspects of phrase structure remain more constant. The upshot is that phrase structure in Japanese is the exact mirror image of phrase structure in English – except for the position of subjects, which is at the beginning of S in both languages, as shown in (44).

(44) (a) English structure



(b) Japanese structure



Other languages with the Japanese style of word order include Hindi, Turkish, and Quechua (the language of the Incas).

Another difference between Japanese and English is that Japanese does not have the question movement transformation. A question word that asks about the object thus typically shows up in the normal position of a direct object in Japanese, which is immediately before the verb (compare the noninterrogative (43)):

(45)	John-ga	dare-o	butta	ka?
	John-NOM	who-ACC	hit	Q
	'Who did John hit?'			

So here, too, we find syntactic diversity – and yet not as much as one might have expected. About 30 percent of languages of the world have question movement to the front of S, whereas 67 percent leave question words in place (Dryer 2005); no other pattern is common. For example, languages in which question words move systematically to the *end* of the clause are very rare, perhaps impossible. So languages apparently have a choice as to whether or not to include the transformation in (14), but that nearly exhausts the choices they have in this domain.

One interesting consequence of this difference between Japanese and English is that it is possible to ask questions in Japanese that are very difficult to ask in English. Question formation involves question movement in English, and this is restricted by conditions like the CED. But since question movement does not happen in Japanese, there is no risk of violating the CED in this language. Thus, (46) is a grammatical question in Japanese, where the *wh*-phrase *dono hon-o* replaces an NP inside a clause that functions as a subject (Lasnik and Saito 1992; see also Huang 1982: 496 on Chinese):

(46)	Mary-ga	John-ga	dono	hon-o	katta	koto-ga
	Mary	John	which	book	bought	fact
	mondai-da	to	omotteru	no ?		
	problem-be	that	think			

'Which book does Mary think that the fact that John bought (it) is a problem?'

This does not mean that the CED is not valid for Japanese; it could be perfectly valid, but simply irrelevant to an example like (46), because (46) does not involve movement. Indeed, relative clause formation in Japanese and Chinese does involve movement; specifically, the head of the relative clause moves out of the sentence rather like English. In this domain, these languages respect the same restrictions that English does (see Huang 1982). So we see a difference between the Japanese and English in terms of exactly what transformations there are, but little or no difference in terms of what conditions must be obeyed by those transformations. This is very much like what we saw in comparing English and Edo in Section 4.1.

4.3 The syntax of Mohawk

My final case study is the native American language Mohawk. English-type languages and Japanese-type languages account for the majority of the languages of the world (82 percent), but not for all of them. The largest remaining group consists of languages like Mohawk, in which no basic word order can be established (Mithun 1987); these are some 14 percent of the overall total. The subject, verb, and object can appear in any logically possible order in Mohawk, with different orders focusing on different elements:

(47)	a.	Sak	ranuhwe's	atya'tawi.
		Sak	likes	dress
		'Sak likes the dress.'		
	b.	Ranuhwe's	Sak	atya'tawi.
		likes	Sak	dress
	c.	Sak	atya'tawi	ranuhwe's.
		Sak	dress	likes
	d.	Ranuhwe's	atya'tawi	Sak.
		like	dress	Sak
	e.	Atya'tawi	ranuhwe's	Sak.
		dress	like	Sak
	f.	Atya'tawi	Sak	ranuhwe's.
		dress	Sak	like

Mohawk thus seems rather radically different from both English/Edo and Japanese: it is not that Mohawk has a slightly different phrase structure; rather it seems to have no fixed phrase structure at all.

Despite this striking difference, it is not hard to see that some of the general principles that underlie phrase structure in English are nevertheless valid in Mohawk as well. For example, we saw that one fact that is encoded in English's phrase structure rules is that all finite clauses must have a subject but not necessarily an object – the Extended Projection Principle (EPP). Something similar is true in Mohawk, but with a twist. Mohawk sentences do not need to have an overt subject (or object) NP; (48) also counts as a complete, well-formed sentence in Mohawk.

(48) Ranuhwe's.
likes
'He likes it.'

But a big part of the reason why this is possible – and why free word order is tolerated in (47) – is because the verb meaning 'likes' (*nuhwe'*) has a prefix on it (*ra-*) that indicates that the subject of the sentence is masculine and singular. Thanks to this prefix, the subject is inferred to be 'he' in (48), and one can recognize that *Sak* (a masculine NP) is the subject in all the variants of (47), rather than 'dress' (which is neuter). All in all, Mohawk has 58 different verb prefixes that express different combinations of subject and object (if any). Taking this into account, we can see that an EPP requirement is valid in Mohawk too, but what it requires is the subject-expressing prefix, rather than a separate independent NP subject. That some kind of subject is required (but not an object) is seen most clearly with weather predicates, which have no inherent semantic need for either a subject or an object. In English, the EPP requires that such predicates take a "dummy" subject pronoun, but not a "dummy" object pronoun, as shown in (49a). In Mohawk there is a parallel requirement that there be a subject-expressing prefix on the verb (here neuter singular *y(o)-*), but there is no requirement that there be an object-expressing prefix, such as -ako.

(49) a. It rained. (Not: **Rained*, **It rained it*, **Rained it*)
b. yo-kvnoru (Not: **kvnor-u*, **y-ako-kvnor-u*, etc.)
it-rains rains it-rains-it

So the phrase structure rule S → NP VP is not universal, but the more abstract principle that helps to shape it – "finite clauses must have subjects" – does apply usefully to Mohawk, and (properly phrased) might be universal.

Another fact about phrase structure in English, Edo, and Japanese is that the direct object combines with the verb to form a unit (the verb phrase) that does not include the subject. A consequence of this is that the object shows up next to the verb in all three – right after it in English and Edo, right before it in Japanese. Example (47) suggests that there is no such difference between subjects and objects in Mohawk: either one can appear closest to the verb. But again, the difference can be seen in a slightly different area of the language. Mohawk has many compound verbs that consist of a noun root followed by a verb root; some examples are shown in (50).

(50) a. Owira'a waha'wahrake' '(the) baby meat-ate'
b. Wa'eksohare' 'She dish-washed.'
c. Wa'kenaktahninu' 'I bed-bought'
d. Wa'khwistatshvri' 'I money-found'

The important point to observe is that in all these examples the noun that is compounded with the verb is interpreted as the *object* of the verb, not as its subject. If one tries to create a compound consisting of a subject and a verb, Mohawk speakers reject it as nonsensical:

(51) *O'wahru wa'kawirake' 'The meat (was) baby-eat(en).'

So Mohawk also has a requirement that the object can combine closely with the verb to make a linguistic unit while the subject cannot. The only difference is in the exact units that are involved: in Mohawk the constraint is seen most clearly when a noun root combines with a verb root to make a compound verb; in English, the constraint is seen when a noun phrase combines with a verb to make a verb phrase (Baker 1988; Baker 1996). Languages like Mohawk in which syntactic conditions seem to apply to complex word formation rather than to phrase-building are known as *polysynthetic* languages.

Consider next transformational and interpretative rules in Mohawk. Question movement happens in Mohawk as in English, although the data looks a bit different because of Mohawk's otherwise free word order. For Mohawk, it is not so surprising that the word "who," inquiring about the object, can appear at the beginning of the sentence as in (52a,b), since noninterrogative objects can appear in this position (see (47e,f)). Rather, what is special about word order in questions in Mohawk is that an order like (52c) is *not* possible, in contrast to nonquestions, in which all word orders are allowed (Baker 1996).

(52) a. Uhka washakokv' ne Sak?
who see NE Sak
'Who did Sak see?'

b. Uhka Sak washakokv'?
who Sak see
'Who did Sak see?'

c. * Sak washakokv' uhka? (etc.)
Sak see who
'Who did Sak see?'

This can be explained by saying that (14) applies in Mohawk as in English. Thus, wherever a question word might start out in Mohawk, it has to move to the front of the relevant sentence, thereby limiting its otherwise free word order.

Mohawk also has the same pronoun indexing rule as English, as shown by the fact that (53) has the same ambiguity as its English counterpart: the subject of the embedded clause may or may not refer to *Sak*, the subject of the main clause.

(55) Sak ihrehre tsi (rauha) vhahninu' ne ka'sere'.
Sak thinks that he will-buy NE car
'Sak thinks that he (=Sak or someone else) will buy the car.'

It is when it comes to the constraints on these rules that the comparison between Mohawk and English becomes most interesting. Mohawk seems to present some complications when it comes to the CED. One consequence of this principle in English is that one can move a question word out of a direct object, but not out of a subject (see (22), (23)). Mohawk is a bit different: it is true that one cannot move a question word out of the understood subject in Mohawk, but one cannot move a question word out of the understood object either. Hence, there is no contrast between (55a) and (55b) in Mohawk, although there is in the English equivalents.

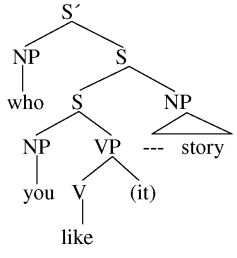
(55) a. *?Uhka wesatsituni' ne akokara'?
who made-you-cry NE story
'Who did the story about make you cry?' (also bad in English)

b. *?Uhka senuhwe's ne akokara'?
who you-like NE story
'Who do you like the story about?' (fairly good in English)

Does this mean that the CED is not valid in Mohawk? Not necessarily! Baker (1996) claims that the CED applies equally to Mohawk and English. The difference between the two languages is not in the condition, but in the phrase structure that the condition applies to. I argued at length that the visible direct object NP is not inside the verb phrase in Mohawk the way that it is in English. We have already seen some elementary evidence that this is so in (47), which shows that the object does not need to be next to the verb in Mohawk, the way it normally is in English. Suppose,

then, that putative "direct objects" in Mohawk are more like adverbial modifiers such as *today* in English; as such they can appear in different places in the sentence, and can be left out altogether (compare *atya'tawi* 'dress' in (47) and (48) with *today* in *Today Chris won*, *Chris won today*, or just *Chris won*). If this is correct, then a "direct object" like [who story] in Mohawk is not inside the VP, but outside it, as shown in (56).

(56)



Given this structure, the CED actually predicts that a question word cannot move out of the "object" NP. The structure is different, but the condition is the same. In fact, "direct objects" can be generated outside the VP in English, too, as long as the normal object position is occupied by a pronoun; this is the so-called right dislocation structure in (57a). Question words cannot be extracted from this sort of dislocated object in English, any more than they can in Mohawk, so (57b) in English is as bad as (55b) in Mohawk.

(57) a. Sak really enjoyed it, the story about Mary.
b. *?Who did Sak really enjoy it, the story about?

Thus, the same condition applies in both languages, initial impressions notwithstanding.

Finally, consider the Disjoint Reference Condition in Mohawk. Again, the patterns look different at first. In English, the DRC implies that a pronominal subject cannot refer to an NP contained in the direct object, although a pronominal object can refer to an NP contained in the subject (see (27), (28)). In contrast, both kinds of coreference are possible in Mohawk:

(58) a.

(Akauha)	wa'akoya'takehnha'	ne	Uwari	akona'tsu.
her	help	NE	Mary	pot

'Mary's pot helped her.' (OK her=Mary)

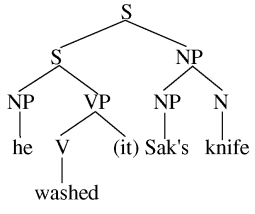
b.

(Rauha)	wahanohare'	ne	Sak	rao'share.
he	wash	NE	Sak	knife

'He washed Sak's knife.' (OK he=Sak)

But here, too, we need to take into account the fact that the phrase structure of Mohawk is a bit different from that of English, in that the overt NP interpreted as the direct object is inside the verb phrase in English, but outside it in Mohawk. Given this, the correct phrase structure for the problematic-seeming (58b) is (59).

(59)



In fact, the subject pronoun does not c-command the NP Sak in (59): the smallest phrase that properly contains 'he' is the smaller S, and this does not contain the NP 'Sak'. Therefore the DRC does not block these two NPs from being coreferent – the correct result. Once again, the condition is the same in Mohawk as in English; the results just come out a bit different because the phrase structure is different. Indeed, coreference between the subject and the possessor of the object is also possible in right dislocation structures in English, as expected:

(60) He_1 finally washed it_2, John_1's knife_2.

The crucial point is that the same assumption about where "direct objects" are located in Mohawk accounts for superficial facts about the variable position of the "object' in the clause, solves a puzzle about the movement of interrogative phrases, and solves a puzzle about coreference possibilities. This convergence is an important sign that this style of analysis is on the right track.

Overall, these three case studies illustrate how languages are thought to differ from each other syntactically, and how they are the same. There are some impressive-looking differences. However, there are also some impressive similarities, and even when languages look most different, many of the differences can be seen as alternative responses to the same underlying constraints. Indeed, the similarities are most striking at the level of syntactic constraints: phrase structure can vary within certain limits, and languages can choose whether to have a particular transformation or not, but the syntactic constraints seem to be the most invariant aspect of syntax.

5 A Glance Ahead

Most of this article has concentrated on the past – on what has been learned from studying syntax in the last 50 years. Some of these matters have been more or less settled for some time. But others reveal questions of ongoing research, which have characterized work of the last 15 years and seem likely to characterize much of the next ten years as well. In this last section, I highlight some of these.

One important question, touched on in Section 2.3, is whether grammatical principles are predominantly general and wide-sweeping, or whether they can be very particular and specific. A closely related question is what to do with apparent exceptions to grammatical principles: how can one acknowledge them in an analysis without giving the whole game away? For example, we saw that English very generally has P-NP order and not NP-P order, whereas it is the other way around in Japanese. But at a finer grain of detail, it has been claimed that English has approximately one P that comes after the NP it is associated with, namely the word *ago* (Culicover 1999).

(61) a. I will go to Paris again *in one year*.
b. I last went to Paris *one year ago*.

Similarly, all heads are at the beginning of their phrases in Edo, except for the one tense marker *n*e, which comes after the verb phrase rather than before it. What should be done about details such as these? Syntacticians are pursuing very different intuitions about this. At the one extreme are the radical construction grammarians, who claim that each "construction" of a language can be encoded explicitly in the grammar: one needs a special rule for *ago* in English, and all other grammatical principles are the same type of thing as this special rule, although they may apply to a larger number of items (Goldberg 1995; Croft 2001). At the other extreme are Chomsky (1981, 1995) and the radical minimalists, who deny that there is any theoretically significant notion of "construction" at all. This implies that the anomalous word order in (61b) should in principle be deducible from the same general laws of phrase structure as (61a). How to resolve issues like this is a topic for ongoing research.

Another current research question is, What is the role of formal and functional explanations in accounting for syntactic phenomena? In the body of this article, I have concentrated on formal explanations: a given sentence is ruled out because it is not *well formed* – it simply cannot be generated by the language's "rules of form" (syntax). For example, a sentence like **Who did a story about – upset you?* is deviant in English because the rule of question movement cannot create it, given the limitation imposed by the CED. The alternative mode of explanation is a functionalist one, which might hold that a sentence like **Who did a story about – upset you?* is not (often) used in English because it does not make much sense or is not very useful (it has no *function*). The fact, mentioned in Section 4.2, that sentences like this are possible in Japanese is prima facie evidence for the formal approach. Presumably it is no more useful or meaningful to ask this sort of question in Japanese than in English, but the sentence is possible because the formal rules are simply different in Japanese (question words do not move in Japanese, so the CED does not apply). But the formal account might not be the whole story. Foley and Van Valin (1984) report the opposite judgment in Lakhota: Lakhota is like Japanese in that question words do not move, yet the following sentence is bad, like its English counterpart.

(62) #Wičhaša wã taku ophethũ ki he wãlaka he.
man a what buy the that you-see Q
'What did you see the man who bought (it)?'

Foley and Van Valin thus naturally draw the opposite conclusion, that the deviance of examples like (62) should be explained functionally rather than formally. And no one doubts that some sentences are unused for functional reasons – example (3b), for one. It is thus a question for ongoing research to tease apart just how these two modes of explanation might work together, and which (if either) is the deeper truth. (See Newmeyer 1998 for general discussion and references.)

As quality syntactic data is rapidly becoming available from more and more languages and dialects, the theoretical question of how languages differ and how they are the same (considered to some extent in Section 4) has become more central. One point of controversy in this domain is whether languages that might look very different on the surface differ from one another in many small ways or in just a few big ways. For example, is word order in Japanese so different from word order in English because many individual phrase structure rules are different, or because one very general phrase structure schema (XP→ X…) is different? The first view is the so-called microparametric approach of Kayne (2005); the second is the so-called macroparametric approach of Baker (1996, 2008). This question is largely the comparative version of the question of whether grammatical analysis should include many small-scale "constructions" or only a few general principles, so comparative syntactic studies that speak to the "many small differences" or "a few large differences" question might bear on that matter too.

A fourth issue of ongoing research has emerged out of thinking further about why English word order is different from Japanese word order. We saw in Section 4.2 that the way words are grouped into phrases in Japanese is virtually identical to the way that words are grouped together in English. Maybe this then is the important truth about syntax per se. Perhaps no linear order is defined at all in the syntax proper, as speakers represent thoughts and sentences in their minds; the mind is, after all, not a linear, serial medium of representation. Perhaps the need to define a word order only arises when it comes time to pronounce that structure, in an effort to communicate it to some other mind. On this view, word order would then be imposed on syntax at the very end, as part of the need to map our multidimensional mental representations of language onto a one-dimensional acoustic stream. If so, the word order of a language might be much less important to its core syntax than the older linguistic tradition has assumed. More generally, there has been a long history of thinking about the constraints that *interpretability* puts on syntactic analysis, about questions such as whether a particular structure makes sense. An area of important growth in the twenty-first century has been thinking more about the constraints

that *pronounceability* might put on the syntax. This has led to many interesting recent proposals (Kayne 1994; Bobaljik 2002; Fox and Pesetsky 2004; among others).

The last cluster of current research questions that I mention here has to do with questions about syntactic categories (parts of speech). How many of these should be distinguished? Do languages always have the same ones? I mentioned above that Japanese is like English in that it has nouns, verbs, adjectives, and prepositions. But does it have articles like *the* and *a*? It seems not to, but some linguists think that such words exist in the syntax of Japanese but are not pronounced (Longobardi 1994, but contrast Chierchia 1998). A related question is whether the position of the complementizer is the same as the position of the inverted auxiliary in English, as I suggested at the end of Section 4.1. There is some evidence that it is: there can be no auxiliary inversion in embedded yes-no questions in English, because these clauses have the special complementizer *whether*:

(63) a. Chris will come.
b. Will Chris come?
c. Pat knows that Chris will come.
d. Pat knows whether Chris will come.
e. *Pat knows whether will Chris come.

It is plausible to say that *will* cannot move in (63e) the way it does in (63b) because the place it would move into is already taken by *whether*. But questions are not the only occasion for subject-auxiliary inversion in English. This inversion also happens when there is a negative word like *never* at the beginning of the sentence:

(64) a. *Never you will see another basketball player like Michael Jordan.
b. Never will you see another basketball player like Michael Jordan.

Now it is possible to have a sequence of a complementizer, followed by a negative word, followed by an auxiliary, followed by subject in English, as shown in (65).

(65) Pat believes that never will we see another basketball player like Michael Jordan.

This detail suggests that the moved auxiliary does not – or does not always – occupy the same position as the complementizer after all; rather they can occupy two subtly different positions. Consideration of facts like these has led many syntacticians over the past 15 years to adopt a very fine-grained approach to syntactic structure – the so-called cartographic approach pioneered by Rizzi (1997) and Cinque (1999). This line of reasoning can lead to what look like extremes, including the idea that every clause in every language is made up of some 50 or more categories, rather than the traditional two or three (verb, tense auxiliary, and complementizer). Is syntactic structure really this fine-grained? Are minor categories like determiner, complementizer, and whatever extra position the word *will* occupies in (65) present in all languages but not pronounced? Or are they only present in the languages in which they can easily be seen? These matters are also being heavily researched and debated at this time.

These, then, are a sample of some of the most important and active areas of research in the last 10 to 15 years, and ones where we look forward to substantial progress in the next 10 to 15 years.

SUGGESTIONS FOR FURTHER READING

Baker (2001) is an accessible, book-length discussion that goes into many of the points made here in more detail.

Newmeyer (1998) is a good source for orientation to some of the controversies surrounding different approaches to the topic of syntax.

Chomsky's own writings are notoriously difficult for beginners, but cannot be avoided entirely. Consider Chomsky (1957) for historical purposes, and Chomsky (1995) for a flavor of more current concerns. (Various textbook presentations of this material are also available.)

Another next step could be surveying the articles in Baltin and Collins (2001).

RELEVANT JOURNALS

Linguistic Inquiry
Natural Language and Linguistic Theory
Syntax

Relevant articles can also be found in the general linguistics journals *Language*, *Lingua*, and *Studia Linguistica*. The journal *Linguistic Typology* overlaps this topic when it comes to matters of comparative syntax and syntactic universals.

NOTES

1. This figure comes from counting VSO languages together with SVO languages.

REFERENCES

Baker, M. (1988). *Incorporation: A Theory of Grammatical Function Changing*. Chicago, IL: University of Chicago Press.

Baker, M. (1996). *The Polysynthesis Parameter*. New York: Oxford University Press.

Baker, M. (2001). *The Atoms of Language*. New York: Basic Books.

Baker, M. (2008). The macroparameter in a microparametric world. In T. Biberauer (ed.), *The Limits of Syntactic Variation* (pp. 351–374). Amsterdam: John Benjamins.

Baltin, M. and Collins, C. (eds) (2001). *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell.

Bobaljik, J. (2002). A-Chains at the PF-Interface: Copies and "Covert" Movement. *Natural Language and Linguistic Theory*, 20, 197–267.

Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics*, 6, 339–405.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1970). Remarks on nominalization. In R. Jacobs and P. Rosenbaum (eds), *Readings in English Transformational Grammar* (pp. 184–221). Waltham, MA: Ginn.

Chomsky, N. (1973). *Conditions on Transformations*: A Festschrift for Morris Halle. New York: Holt, Rinehart and Winston.

Chomsky, N. (1975). *Logical Structure of Linguistic Theory*. Chicago, IL: University of Chicago Press.

Chomsky, N. (1977). *On Wh-Movement*: Formal Syntax. New York: Academic Press.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. and Lasnik, H. (1977). Filters and control. *Linguistic Inquiry*, 11, 1–46.

Cinque, G. (1999). *Adverbs and Functional Heads: A Cross-linguistic Perspective*. New York: Oxford.

Croft, W. (2001). *Radical Construction Grammar*. Oxford: Oxford University Press.

Culicover, P. (1999). *Syntactic Nuts: Hard Cases, Syntactic Theory, and Language Acquisition*. New York: Oxford University Press.

Dryer, M. (2005). Order of subject, object, and verb. In M. Haspelmath, M. Dryer, D. Gil, and B. Comrie (eds), *The World Atlas of Language Structures* (pp. 386–89). New York: Oxford University Press.

Foley, W. and Van Valin, R. (1984). *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.

Fox, D. and Pesetsky, D. (2004). Cyclic linearization of syntactic structure. *Theoretical Linguistics*, 31, 1–46.

Gazdar, G., Klein, E., Pullum, G., and Sag, I. (1985). *Generalized Phrase Structure Grammar*. Cambridge, MA: Harvard University Press.

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: University of Chicago Press.

Grimshaw, J. (1997). Projections, heads, and optimality. *Linguistic Inquiry*, 28, 373–422.

Huang, C.-T. James (1982). *Logical Relations in Chinese and the Theory of Grammar*. PhD disseration, MIT.

Jackendoff, R. (1977). *X-bar Syntax*. Cambridge, MA: MIT Press.

Kayne, R. (1994). *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Kayne, R. (2005). Some notes on comparative syntax, with special reference to English and French. In G. Cinque and R. Kayne (eds), *The Oxford Handbook of Comparative Syntax* (pp. 3–69). New York: Oxford University Press.

Koopman, H. (1982). Control from Comp and comparative syntax. *The Linguistic Review*, 2, 365–81.

Kuno, S. (1973). *The Structure of the Japanese Language*. Cambridge, MA: MIT Press.

Lasnik, H. (1989). *Essays on Anaphora*. Dordrecht: Kluwer.

Lasnik, H. and Saito, M. (1992). *Move Alpha: Conditions on its Applications and Output*. Cambridge, MA: MIT Press.

Longobardi, G. (1994). Reference and Proper Nouns. *Linguistic Inquiry*, 25, 609–66.

Mithun, M. (1987). Is basic word order universal? In R. Tomlin (ed.), *Coherence and Grounding in Discourse* (pp. 281–328). Amsterdam: Benjamins.

Newmeyer, F. (1998). *Language Form and Language Function*. Cambridge, MA: MIT Press.

Reinhart, T. (1983). *Anaphora and Semantic Interpretation*. Chicago, IL: University of Chicago Press.

Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (ed.), *Elements of Grammar* (pp. 281–338). Dordrecht: Kluwer.

Ross, J. R. (1967). *Constraints on Variables in Syntax*. PhD dissertation, MIT. (Published as *Infinite Syntax!* (1986), Norwood, NJ: Ablex).

Stewart, O. (2001). *The Serial Verb Construction Parameter*. New York: Garland.

14 Formal Semantics*

SHALOM LAPPIN

1 Introduction

When people talk, they generally talk about things, events, and situations in the world. They are able to do this because they represent connections between the expressions of their language and extra-linguistic phenomena in a fully systematic way. The meaning of a sentence in a language is, to a large extent, dependent upon the ways in which the words and phrases from which it is constructed can be related to situations in the world. Speakers of a language are able to communicate effectively with each other because they have internalized the same rules for pairing the lexical items of the language with nonlinguistic elements, and they use the same procedures for computing the meaning of a syntactically complex phrase from the meanings of its parts. Therefore, speakers will, in general, converge on the same sets of possible language–world connections which they assign to the sentences in their discourse. Formal semanticists seek to understand this aspect of linguistic meaning by constructing precise mathematical models of the principles that speakers use to define those relations between expressions in a natural language and the world which support meaningful discourse.[1]

Consider an example. Assume that two students in a class are discussing the class's progress on a term paper. One student asserts (1a) and the second responds with (1b).

(1) a. John has finished his paper.
b. No one in the class has finished his / her paper.

For the second speaker to understand (1a), he / she must be able to pick out the person corresponding to *John*. He / She must also know what property *finished his paper* expresses and recognize that the first speaker is claiming that the person corresponding to *John* has this property. If (1b) is true, then it implies that (1a) is false by virtue of the fact that (1b) states that no person in the class has the property of having finished his / her paper. Therefore, assuming that the second speaker understands both (1a) and (1b), then he / she recognizes that asserting (1b) involves making a statement which is incompatible with the one made by the first speaker.

To competent speakers of English all of this is thoroughly obvious. This is because we have already internalized the semantics of English, which we rely on in understanding the partial

* I am grateful to Ruth Kempson, Gabriel Segal, and the editors of this volume for helpful comments on earlier drafts of this paper. I am solely responsible for any shortcomings which remain.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

(and informal) description of the semantic competence required to interpret the simple dialogue in (1). But consider what is involved in developing a complete theory of this semantic competence which renders formal and explicit our tacit knowledge of linguistic meaning rather than presupposing it. Such a theory will specify what sort of properties verb phrases (VPs) like *finished his paper* refer to, and it will model properties in general as formal objects of a kind which can apply to the entities of the sort identified by *John*. It will also capture the important semantic distinctions and similarities between proper names like *John* and quantified noun phrases (NPs) such as *no one*. Specifically, while *John* selects an individual, *no one* does not. On the other hand, both kinds of expression can combine with the predicate *finished his paper* to yield a meaningful statement. It is also necessary to explain the difference in the anaphoric relation which holds between the pronoun *his* (/ *her*) and the subject NP in (1a) and (1b).

A complete semantic theory will apply not only to the sentences in (1), but to all syntactically well-formed sentences of the language. Specifically, it must explain our capacity to assign interpretations to an unbounded number of grammatical sentences. Given that we can only represent a finite number of primitive semantic elements, this capacity requires the recursive application of rules to the meanings of expressions in order to derive interpretations for larger phrases.[2] There is, then, a direct formal analogy between the syntactic component of the grammar, which employs recursive procedures to generate a (potentially) infinite set of sentences from smaller lexical and phrasal units, and the semantics, which combines the meanings of these units into the interpretations of the sentences in which they are contained.

In the following sections I will look at some of the central questions which arise in constructing a formal semantic theory for natural language, and I will briefly indicate several of the major lines of research which formal semanticists have pursued in their attempts to answer these questions.

2 Meanings and Denotations

Semanticists have traditionally focused on theories of meaning which apply to sentences that make statements, and are taken to be either true or false. The assumption underlying this approach is that this type of sentence provides a paradigm of the sort of relationship between linguistic expressions and the world which is at the core of linguistic meaning. An additional assumption is that if it is possible to construct a successful account of the meaning of declarative sentences used to make statements, then this account can be generalized to nondeclarative sentences, like interrogatives that are employed for asking questions, and imperatives which communicate commands.[3]

It is possible to locate the beginnings of modern formal semantics in the work of the German logician, Frege, who created the foundations of first-order logic.[4] We have identified one of the key tasks of a semantic theory as the specification of a systematic correspondence between categories of expressions in a language and types of entities in the world. The main syntactic categories which Frege identifies in natural language correspond to the types of first-order logic. These types are (i) individual terms (names of individuals, and variables that occur in the same positions in sentences that names do), (ii) predicates (terms for properties and relations), (iii) connectives (*and, or, if … then*, and *not*) for building composite sentences and negations out of component sentences, and (iv) quantifiers that are linked to variables (bind the variables). Proper names, like *John*, and definite descriptions like *the prime minister* are treated as individual terms that occupy the positions of arguments in predicate terms. VPs like *sings* and *introduced the bill* are one-place predicates in that they apply to single arguments to yield statements.

Frege claims that for each logical type an expression of that type can take a certain sort of entity as its *denotation* (the thing that it stands for). Individual terms denote individuals in the world (more precisely, in the domain of discourse shared by the speaker and his / her hearers). If

one knows how a declarative sentence like (2a, b) stands in relation to the world, then one knows whether it is true or false.

(2) a. John sings.
b. The prime minister introduced the bill.

In this sense, the primary semantic value of a declarative sentence is its truth or falsity, and Frege takes declarative sentences to denote truth-values. One-place predicates denote functions from individuals to the truth-values true or false. Every function is a mapping from a set of arguments (its domain) to a set of values (its range). Therefore, the function *f* which a one-place predicate denotes can be represented as the set of objects in *f*'s domain for which *f* yields the value true. The VP *sings*, for example, denotes the set of things in the domain of discourse of which *sings* is true. (2a) is true if and only if (iff) the individual which *John* denotes is an element of the set that *sings* denotes, and similarly for (2b).

This schema for category–type correspondence extends naturally to sentences formed with logical connectives like *and, or, if … then*, and negation, as in (3).

(3) a. John sings and Mary dances.
b. John sings or Mary dances.
c. If John sings then Mary dances.
d. John doesn't sing.

Two-place connectives denote functions from pairs of truth-values to a truth-value. So *and* maps two true sentences into the value true, and every other combination of truth-values into false. (3a) is true iff both *John sings* and *Mary dances* are true. *Or* maps any two false sentences into the value false, and any other pair of values into true. (3b) is true iff at least one of the disjuncts connected by *or* is true. *If … then* is false if the antecedent (the sentence immediately following *if*) is true and the consequent is false, and true otherwise. It follows that (3c) is true iff either *John sings* is false or *Mary dances* is true. Finally, a negated sentence is true iff the sentence to which the negation applies is false. (3d) is true iff *John sings* is false.

What about quantified NPs like *no one* in (1b), and the subjects of (4)?

(4) a. Someone sings.
b. Everyone dances.

Unlike individual terms, they do not denote individuals in the domain, but they do seem to occupy the same grammatical category as these terms. How, then, do we interpret them? Frege revolutionized logic by treating quantifiers as second-order functions, or, equivalently, second-order property (set) terms (see note 4 for the distinction between first- and second-order terms). On this view, (1b) and (4a, b) are not statements in which a predicate is applied to an argument, but quantified sentences in which a term that corresponds to a property of a set applies to a predicate (a term that denotes a set). (4a) is true iff the set of things that sing has at least one element, and (4b) is true iff everything in the domain of people dances. (1b) is equivalent to (5).

(5) It is not the case that someone in the class has finished his / her paper.

This sentence is true iff the set of people in the class who have finished their respective papers is empty.

First-order logic has two basic quantifiers, *every* and *some*. Each of these quantifiers can be expressed as an operator that is prefixed to a sentence and associated with variables which appear in the argument positions of predicates in the sentence. The symbol commonly used for *some* is ∃x (*for some x*), and for *every* it is ∀x (*for every x*). The symbols frequently used for negation, conjunction, and implication are ~ (*it is not the case that*), & (*and*), and → (*if … then*), respectively. Let's

substitute *some student* for *someone* in (5) in order to give explicit expression to the restriction of the quantifier *some* to the set of students in the domain. Then we can represent (5) in first-order logic as (6a), which is equivalent to (6b).

(6) a. ~∃x(student(x) & finished x's paper(x))
b. ∀x(student(x) → ~ finished x's paper(x))

(6a) states that it is not the case that there is an object x in the domain which is both a student and finished its own (x's) paper. (6b) states that for every x in the domain, if x is a student, then it is not the case that x finished x's paper. Notice that each occurrence of the variable *x* is interpreted relative to the quantifier prefixed to the sentence where the variable appears. The quantifiers ∃x and ∀x *bind* the variable *x* in (6a) and (6b), respectively.

On Frege's view individual terms and variables, unlike quantifiers, are arguments of predicates. Therefore, (1a) is expressed in first-order logic by (7).

(7) finished john's paper (john)

Notice that the anaphoric relation between the pronoun *his* and *John* in (1a) is captured by substituting the denotation of *John* for the pronoun in (7). By contrast, the anaphoric dependence of *his* (*her*) upon its quantified NP antecedent *no one* in (1b) is represented by using a bound variable for the pronoun in (6a, b).

Definite descriptions pose an interesting problem for a theory which attempts to explain the meaning of an expression in terms of its denotation. The definite descriptions *the former governor of Arkansas* and *the forty-second president of the United States* denote the same object, Bill Clinton. Therefore, if we substitute one for the other as the argument of the predicate *plays the saxophone*, the truth-value of the resulting statement should not be affected. In fact, (8a) and (8b) do have the same truth-values.

(8) a. The former governor of Arkansas plays the saxophone.
b. The forty-second president of the United States plays the saxophone.

However, the two descriptions do not have the same meaning, and (8a) and (8b) assert different statements. *The former governor of Arkansas* identifies the person who was the previous governor of Arkansas, but who no longer holds this position, and *the forty-second president of the United States* denotes the individual who was the current president at a particular point in time. The difference in meaning can be brought out clearly by evaluating (8a, b) relative to a particular point in time. During the 1992 American presidential election campaign, (8a) was true, as Clinton was the former governor of Arkansas, but not yet the president. (8b), however, was false, because George Bush was the president.

The observation that the denotation of an expression does not exhaust its meaning led Frege to factor meaning into the two components of *denotation* and *sense*. He characterizes the sense of an expression as the principle for determining its denotation. Therefore, two terms with the same sense will always have identical denotations, but, as (8) indicates, the converse does not hold. Frege does not give a precise description of the formal entities which correspond to senses. Carnap (1947) substitutes *extensions* for denotations and *intensions* for senses. Extensions correspond closely to Frege's denotations. We can take the extension of an expression E to be the entity which it denotes, where this entity is of the kind appropriate for E's logical type. The extension of a declarative sentence is its truth-value, of a name an individual object, and of a predicate a set of objects (or, in the case of a relation, a sequence of objects).

The intension of an expression E is essentially a rule for identifying E's extension in different situations. Carnap characterizes intensions as functions from possible worlds to denotations, where a possible world can be thought of as the result of specifying the properties and relations which hold for the objects of a domain in a way that defines a complete state of affairs for the

entities of the domain. The actual world is one of many (in fact, an infinite number of) possible worlds. The intension of an expression takes a possible world as an argument and yields the extension of the expression in that world as its value. Therefore, the intensions of *the former governor of Arkansas* and *the president of the United States* identify (i) the person who satisfies the property of being the previous governor of Arkansas and (ii) the person who is currently the president of the United States, respectively, in each world. These two denotations converged on the same individual in the actual world, but are distinct in other possible worlds (and times). Similarly, the intension of the VP *plays the saxophone* picks out the set of objects which play the saxophone for each world. The intension of a sentence assigns it a truth-value in each possible world. We obtained (8b) from (8a) by substituting one description for another with the same extension but a distinct intension. The substitution produced a sentence with the same extension (truth-value) in the actual world (at the time), but a different intension (proposition).

We observed that one of the main tasks of semantic theory is to explain how speakers compute the meanings of complex phrases from the meanings of their parts. Frege adopts the principle of *compositionality* as a condition of adequacy on any account of meaning. Compositionality requires that the meaning of any well-formed phrase in a language be a function of the meanings of its syntactic components. This condition implies that, for any phrase P, given the meanings of the constituents of P, there is a function which maps these meanings into the meaning of P. This principle has enjoyed wide acceptance throughout the history of semantic theory. Clearly, if an account of meaning satisfies compositionality, it specifies the way in which the interpretations of complex structures are generated from their constituents. However, as we will see in Section 5, it is possible to construct noncompositional semantic theories which also fulfill this task.

On the Frege–Carnap approach, the principle of compositionality yields two distinct subprinciples: (i) the extension of a phrase is a function of the extensions of its parts; (ii) the intension of a phrase is a function of the intensions of its parts; truth functional connectives produce complex sentences that satisfy (ii). So, for example, the truth-value of (3a) is a function of the truth-value of the two conjuncts of *and*.

(3) a. John sings and Mary dances.

However, verbs like *believe*, which map propositions into properties (sets) of individuals are problematic. Unlike truth functional connectives, *believe* is sensitive to the intension as well as the extension of the sentence which it takes as its grammatical complement. Substituting one complement sentence for another with the same truth-value but a different proposition can alter the extension, as well as the intension of the entire VP.

In addition to the Frege–Carnap view there is another approach, which dispenses with intensions and seeks to construct a theory of meaning solely in terms of the contributions which expressions make to the truth (i.e., extension) conditions of sentences. This approach is developed by Davidson, and it takes as its starting point Tarski's (1933) definition of truth for first-order languages.[5] Tarski constructs a recursive definition of the predicate *true-in-L* for a class of first-order languages similar to the first-order language characterized by Frege. The definition proceeds stepwise first to elementary sentences constructed from individual terms (constants or names, and variables) and predicates, next to compound sentences formed by applying truth functions to other sentences, and finally to quantified sentences. For each sentence S of type T in language L, it specifies the truth conditions for S in terms of the relations which must hold among the denotations of the constituents of S. As a result, Tarski's truth definition generates appropriate truth conditions for the full set of well-formed sentences of L.

Davidson regards Tarski's truth definition as the paradigm of a semantic theory.[6] If to know the meaning of a declarative sentence is to know its truth conditions, then Tarski's definition gives an explanation of sentence meaning in terms of a precise and fully systematic account of the connections between sentences and the world. It does this in a way which exhibits how the interpretations of sentences are built up from the interpretations of their constituents.[7] Davidson's

general strategy is to associate the sentences of a natural language with first-order logical forms to which a Tarskian truth definition can apply.

Frege and Carnap on one hand, and Davidson on the other, share the assumption that the sentences of natural language are analyzed in terms of the types of first-order languages, specifically, individual terms, k-place predicates (predicates that take k number of arguments), truth-functional connectives, and first-order quantifiers like *every* and *some*. Montague (1974) discards this assumption, and establishes a far richer and more expressive type system for intensional semantics.[8]

The basic framework which Montague adopts for developing a formal syntax and semantics for natural language is categorial grammar.[9] In this system a small number of syntactic categories are taken as basic. All other categories are functions from input expressions of a certain category to output expressions of a given category. Assume, for example, that we take sentences and expressions which denote individuals (i.e., names) as basic, and that we indicate the former category by *t* (for truth-value) and the latter category by *e* (for entity). Categorial grammarians represent functional categories as slashed expressions in which the argument term appears to the right of the slash and the output term is to the left. A VP and a common noun are both a t/e (a function from names to sentences), a transitive verb is a (t/e) / e (a function from names to VPs), a verb like believe, which takes a sentential complement, is (t/e) / t (a function from sentences to VP's), an NP is a t / (t/e) (a function from VP's to sentences), and a determiner is a (t/(t/e)) / (t/e) (a function from common nouns to NPs. In each case, a slashed category expression combines with a term of the kind indicated to the right of the slash in order to produce a term of the sort which appears to the left of the slash.

Consider the sentences in (9a, b).

(9) a. Mary sings.
b. John likes Mary.

If we take *Mary* as name of category e, then *sings*, which is an intransitive verb of type e/t combines with the e term *Mary* on its left to produce a t term (sentence). Similarly, the transitive verb *likes* in (9b) is of category (t/e) / e. It combines with the e term object *John* on its left to yield an intransitive verb (VP) *likes John* of type e/t. This e/t term takes the e term *John* on its left to give a t term as its value.

Montague establishes a strict correspondence between the syntactic categories and semantic types (denotation types) of the grammar. The correspondence is expressed as a homomorphism, which is a mapping that assigns a single semantic type to each syntactic category. Sentences denote truth-values, and predicates (VPs and common nouns) denote functions from individuals to truth-values (equivalently, sets of individuals). For all other categories where *f* is a syntactic function of the form *a/b*, the semantic value (denotation) of *f* will be a function from the intension of *b* (*f*'s argument) to the extension of *a* (*f*'s value). So, for example, *believe* is an element of the category of functions from sentences to VPs, and it denotes a function from sentence intensions (propositions) to sets of individuals. This set contains the people who stand in the belief relation to the proposition expressed by the complement of *believe*. Montague grammar defines the category-type correspondence recursively for every expression of the language in a way which satisfies the principle of compositionality. Therefore, the meaning of every phrase in the language is a function of the meanings of its parts. Moreover, given the functional nature of syntactic categories and semantic types, it is possible to generate as many of each as one requires to accommodate complex syntactic structures in natural language. Each functional category will always map into a corresponding semantic type that specifies the set of possible denotations for the expression. Although there are, in principle, an unbounded number of functional categories and types, only a finite (and fairly small) number are used in the grammar of a language.

Two important differences between the Montague and Davidsonian approaches concern (a) the analysis of modification and (b) the treatment of NPs. Consider modifiers of common nouns, like the adjective *green* in *green house*, and modifiers of VPs, like the temporal adverb *on Thursday* in *arrived on Thursday*. On the Davidsonian view, modifiers are predicates which apply to individuals. A modified common noun is taken to be the conjunction of several predicates. (10a), for example, is analyzed as (10b), which states that there is an object x such that x is a house, x is green, and Mary has x.

(10) a. Mary has a green house.
b. ∃x(house(x) & green(x), & has(mary, x)).

Adverbs are also taken as predicates, and they are applied to events, which are included in the domain of entities.[10] (11a) is interpreted as (11b), which asserts that there is an event e that has the property of John arriving in e, and e occurred on Thursday.

(11) a. John arrived on Thursday.
b. ∃e(arrived(j,e) & on_Thursday(e)).

For Montague both common nouns and VPs are predicates. Syntactically, modifiers are functions from predicates to predicates, and semantically they are functions from predicate intensions to predicate extensions (sets). In (10a) *green* denotes a function which takes the intension of *house* as its argument and yields the set of green houses as its value. Similarly, in (11a) the function which *on Thursday* denotes applies to the intension of *arrived* to give the set of things that arrive on Thursday.

Davidson's account is attractively simple and straightforward. It reduces all modification to first-order predication. However, it encounters two problems. First, it assigns a semantic type to modifiers which is quite remote from their syntactic role. Syntactically modifiers are functions that apply to expressions to produce expressions of the same category. Adjectives and relative clauses apply to nouns to create modified nouns, and adverbs apply to VPs to create modified VPs. However, Davidson's analysis treats modifiers as semantic predicates that have the same kind of denotation as the predicates they modify. So, for example, in (10b) both the noun *house* and its modifier *green* are taken as one-place predicates. Similarly, in (11b) the verb *arrived* corresponds to a two-place predicate, and its adverb *on Thursday* is analyzed as a one-place predicate.

Second, the analysis does not extend to modifiers that produce expressions whose meanings cannot be taken as the conjunction of two predicate extensions. The adjective *toy* and the adverb *allegedly* in (12a, b), respectively, are examples of such nonextensional modifiers.

(12) a. John has a toy car.
b. Mary allegedly submitted her paper.

(12a) cannot be paraphrased as there is an x such that x is a toy, x is a car, and John has x. The sentence means that John has an object which resembles a car in certain respects, but which is not a car. Similarly, (12b) cannot be taken to assert that there is an event e in which Mary submitted her paper, and e allegedly occurred. If (12b) is true, then there may have been no event of Mary submitting her paper. Nonextensional modifiers require a different kind of semantic representation. They cannot be analyzed as predicates that apply to objects and events. Therefore, Davidson's approach does not provide a unified treatment of modification.

Montague's account avoids both difficulties. The semantic type of a modifier is a function which works in strict parallelism with its syntactic function. Syntactically it is a function from predicates to predicates, and semantically it denotes a function from the intension of its syntactic argument to the extension of its syntactic value. An adjective denotes a function from the intension (property) of the noun to which it applies to the set of objects that the modified noun

denotes. An adverb has as its denotation a function from the intension (property) of the VP which it modifies to the set of objects that provide the extension of the modified VP. This account covers nonextensional modifiers by virtue of the fact that the function that a modifier denotes applies to predicate intensions rather than extensions. In (12a) the denotation of *toy* applies to the intension of *car* rather than the set of cars (the extension of *car*) to give the set of toy cars (not the set of things which are both toys and cars). In (12b), the denotation of *allegedly* takes the intension of *submitted her paper*, not the set of submitted papers as its argument. It yields the set of (female) things which allegedly submitted their respective papers as the extension of the modified VP.

The disadvantage of Montague's treatment of modification is that it does not express the fact that when an extensional modifier applies to a predicate, it does produce a predicate whose interpretation is equivalent to the conjunction of two predicates. In order to capture this property of modification, it is necessary to add a set of rules to the semantic part of the grammar which insure that (10a) implies that Mary has a house, and it is green, and (11a) implies that John arrived and his arrival was on Thursday. Therefore, while Montague's approach offers a unified account of modification, it does so at the cost of a more complicated treatment of extensional modifiers.

Turning to the interpretation of NPs, we have already observed that Davidson follows Frege in taking proper names to be terms that denote individuals and appear as arguments of predicates, while analyzing quantified NPs as operators which bind variables in argument positions. Therefore, (13a) and (14a) are assigned the logical forms in (13b) and (14b), respectively.

(13) a. John sings
b. sings(john)

(14) a. Every student sings.
b. $\forall x(\text{student}(x) \rightarrow \text{sings}(x))$

The advantage of this view is that it associates sentences like (13a) and (14a) with first-order formulas for which a Tarskian truth definition is available. The semantic intepretation of the sentence follows directly from its logical form. Notice, however, that while names and quantified NPs appear in the same syntactic roles (subject, object, indirect object, object of a preposition, etc.), they are mapped into distinct semantic types.

Because names and quantified NPs occupy the same syntactic roles (subject, object, object of a preposition, etc.), Montague treats them as members of a single syntactic category. He characterizes them as functions which take VPs as arguments to produce sentences (i.e., they are functions of the sort t / (e/t)). Recall that all elements of a given category receive the same semantic type in accordance with the general principle that specifies the category–type correspondence. It follows from this principle that all NPs denote functions from VP (predicate) intensions to truth-values. Predicate intensions are properties of individuals, and, as we have observed, a function from entities to truth-values is equivalent to the set of those entities to which it assigns the value true. Therefore, the function which an NP denotes can be represented by a set of properties (the set of properties for which it gives the value true). Recall that Frege treats quantifiers as second-order properties, i.e., as sets of sets. If we simplify Montague's account slightly by taking NPs as functions from predicate extensions (sets), rather than predicate intensions, to truth-values, then NPs denote sets of sets. For Montague, all NPs are, in effect, quantifiers. This semantic type is referred to as the class of *generalized quantifiers* (GQs), where a GQ is a set of sets of individuals.[11]

It is clear how an NP like *every student* can be interpreted as a generalized quantifier. It denotes the set of sets (or properties) each of which contains (at least) every student. (14a) is true iff the set of things that sings is an element of this set of sets. The set of singers is an element of the set of sets denoted by *every student* iff the set of singers contains the set of students as a subset, which is equivalent to the assertion that every thing which is a student sings. The truth conditions that Montague's GQ analysis assigns to (14a) are equivalent to those of the first-order sentence in (14b).

But it is not so obvious how proper names can be accommodated in this system. Montague's solution to this problem is to treat a name as denoting not an individual, but the set of sets containing an individual (the property set of an object). (13a) is true, then, iff the set of singers is an element of the set of sets containing John, which holds iff John is an element of the set of singers. As there is a one-to-one correspondence between the property set of an individual and the individual itself, these truth conditions reduce directly to those for (13b).

The GQ analysis sustains a uniform semantic representation of NPs. However, it does so at the price of certain complications. These become particularly clear in the case of NPs in nonsubject position, like *Mary* in (15a) and *every paper* in (15b).

(15) a. John likes Mary.
b. Max read every paper.

The truth conditions of these sentences can be expressed by the first-order sentences in (16), where *likes* and *read* are naturally represented as denoting relations between individuals.

(16) a. likes(john, mary)
b. $\forall x(\text{paper}(x) \rightarrow \text{read}(\text{max}, x))$

However, if *Mary* denotes a GQ, then *likes* denotes a function from GQs to sets.[12] This function must be characterized as applying to Mary's property set to yield the set of objects that like Mary as the denotation of *likes Mary*. Similarly, *read* in (15b) maps the GQ denoted by *every paper* into the set of objects which read every paper. Therefore, we are forced to adopt the counterintuitive idea that transitive verbs stand for relations between individuals and sets of sets (GQs) rather than the more natural view that they denote relations between individuals.

An important advantage of the GQ approach is that it covers NPs like *most students*, which cannot be reduced to restricted first-order quantifiers like *every / some student*. To see this, consider what sort of logical form would correspond to (17).

(17) Most students sing.

Assume that *most*(x) is a variable binding operator like $\exists x$ and $\forall x$, and that C is a truth-functional connective. Then the logical form for (17) will be an instance of the schema (18), with an appropriate connective substituted for C.

(18) most(x)(student(x) C sings(x))

But there is no truth-functional connective which can be substituted for C to yield a first-order sentence with the correct truth conditions for (17). The reason for this is that *most*(x) quantifies over the entire domain of objects, while in (17) the natural language determiner *most* expresses a relation between the set of students and the set of singers which cannot be captured by a truth-functional connective. If we use & for C, then (18) states that most objects in the domain are both students and singers. Alternatively, if we take C to be $\rightarrow$, then (18) asserts that for most objects x, if x is a student, then x sings. (17) does not make either of these claims. It states that the majority of objects in the set of students are singers. In fact, there is no first-order sentence whose truth conditions give the intended interpretation of (17).[13]

Taken as a GQ *most students* denotes the set of sets which contain more than half the set of students. (17) is true iff the set of singers is in this set. This condition holds iff the the number of students who sing is greater than half the number of students. Clearly, these are the correct truth conditions for (17). The existence of quantified NPs like *most students* shows that the meanings of some expressions in our language cannot be fully expressed in terms of the truth conditions of first-order sentences, and it is necessary to use more powerful systems, like GQ theory, to model the semantics of natural language.

3 Dynamic Semantics: Beyond Static Sentence Meanings

Until now we have been concerned with the interpretation of sentences as static and independent units of meaning. This perspective allows us to focus on the way in which the meanings of a sentence's constituents contribute to its truth conditions. But, in fact, we generally encounter a sentence as a part of a discourse, where we understand it on the basis of preceding contributions to the conversation. When we situate sentence meanings in a discourse, they are no longer static objects, but active devices that have the capacity to inherit semantic content from previous sentences, modify it, and pass on the new information to the next sentence in the sequence.

The simple two-sentence discourse in (19) illustrates this dynamic aspect of meaning:

(19) John came in. He sat down.

We understand *he* in the second sentence as referentially dependent upon *John* in the first. We also impose an ordering relation on the events described by these sentences, so that we take John to have sat down after he entered. The interpretation of *He sat down* depends upon the information introduced by *John came in*.

Now consider the discourse in (20).

(20) A man came in. He sat down.

Although it resembles (19), there is an important difference. The proper name *John* denotes an individual, but the indefinite NP *a man* does not. Notice also that because the pronoun occurs in a different main clause than the indefinite, we cannot treat it as a variable bound by an existential quantifier. In general, pronouns can only be interpreted as bound by a quantifier in the same clause. In (21a), *his* can be understood as a variable bound by the quantifier corresponding to *every boy*, as in (21b).

(21) a. Every boy handed in his paper.
b. $\forall$x(boy(x) $\rightarrow$ handed in x's paper(x))
(For every x, if x is a boy, then x handed in x's paper.)
c. Every boy arrived. He had a good time.
d. $\forall$x(boy(x) $\rightarrow$ arrived(x)). had_a_good_time(x)
(For every x, if x is a boy, then x arrived. x had a good time.)

However, such an interpretation is not available for *he* in (21c). The quantifier in (21d) cannot bind the variable *x* in the following sentence, which is out of its scope. Therefore, *x* is free (unbound by the quantifier) in the second sentence of (21d). This sentence says that x had a good time without placing any restrictions on the values of *x*. We could have used *y* instead of *x* in the second sentence of (21d), which would give *had_a_good_time*(*y*), without changing the meaning of (21d).

The interpretation of *A man came in* in (20) makes available a possible referent which *he* can be used to identify in the next sentence. However, it is not clear precisely which part of the meanings of these two sentences creates this entity.

The cases in (22) provide examples of a similar but more complex anaphoric relation between a pronoun and an indefinite NP.[14]

(22) a. Every man who owns a donkey beats it.
b. If a man owns a donkey, he beats it.

As with the pronoun in (20), *it* is not within the scope of its antecedent, the indefinite NP *a donkey*, in either (22a) or (22b). This NP is contained either in a relative or subordinate clause rather than in the main clause where *it* appears. *He* is not within the scope of *a man* in (22b) for the same

reason. However, both pronouns appear to function like variables bound by the universal quantifier *every*. On their most natural readings, (22a, b) assert that for every man x and every donkey y, if x owns y, then x beats y. The quantified NP subject in (22a) and the antecedent *if* clause in (22b) give rise to the representation of a set of ordered pairs <a, b> such that a is a man, b is a donkey, and a owns b. For each such pair, *he* in (22a) identifies a, and *it* in both sentences selects b. The problem is that because the pronoun *it* is anaphorically dependent upon the indefinite *a donkey* in (22a, b) it does not correspond to a variable bound by a universal quantifier. It is not obvious, then, how it is possible to interpret (22a, b) as equivalent in truth conditions to a sentence in which *it* is bound by a universal quantifier corresponding to *every donkey*.

There are three main approaches to dynamic anaphora, and I will briefly sketch each one in turn. The first is discourse representation theory (DRT).[15] In this framework an indefinite NP is treated not as a quantified NP, but as an expression which introduces a discourse referent that satisfies the content of the indefinite description. In (20) *a man* introduces an object u, which satisfies the predicate *man*, into the store of information available within the discourse. The sentence also applies the predicate *came in* to u. Therefore, the first sentence of (20) adds the conditions *man*(u) and *came in*(u) to the discourse information store. As u is now accessible at future points in the discourse, it is possible to use a suitable pronoun to refer to it. The second sentence of (20) contributes the condition *sat down*(u), which is obtained by taking u as the value of *he*. The conjunction of these conditions on u yields a discourse representation structure that holds iff there is a man who came in and that man sat down, which is the desired reading of the sequence.

Applying this approach to (22b), the two indefinite NPs in the antecedent clause introduce two distinct discourse referents u and v, and the conditions *man*(u), *donkey*(v), and *owns*(u,v). These referents and conditions are accessible to the consequent clause, where u and v are substituted for *he* and *it*, respectively, to produce the condition *beats*(u,v). However, the relation between the two clauses is not that of a simple sequential conjunction, as in (20), but a conditional connective. Therefore, it is necessary to interpret the combined discourse structure as asserting an *if … then* relation between the conditions of the antecedent and that of the consequent. On the preferred reading of (22b), the conditional sentence is within the scope of the implied adverb of universal quantification *in every case* (or *always*). Applying this quantifier to the conditional discourse structure gives a set of conditions that hold iff for every case, if there is a pair containing a man and a donkey which he owns, then the first element of the pair beats the second. This is the required interpretation for (22b). Assume that the universal quantifier *every* of *every man who owns a donkey* in (22a) sets up a universal conditional relation between the conditions imposed by the modified noun *man who owns a donkey* and those of the VP *beats it*, and that it also introduces a variable x into both sets of conditions. The antecedent of this conditional contains *man*(x), *donkey*(u), and *owns*(x,u), and the consequent adds *beats*(x,u). This discourse representation structure specifies the same interpretation as the one for (22b).

The second approach to dynamic anaphora is the dynamic binding account.[16] It retains the traditional view of indefinites as existentially quantified NPs. In addition to the classical logical connectives and quantifiers it introduces dynamic counterpart operators whose scopes can extend beyond single clauses. The dynamic existential quantifier $\exists^{d}x$ has the effect of introducing a discourse referent associated with the variable x which can be inherited by the informational state (discourse model) that serves as the input to a subsequent sentence. The dynamic conjunction $\&^{d}$ passes the referents in the information state produced by its first conjunct to the interpretation of the second. These dynamic operators are used to represent (20) as (23a), where the dynamic existential quantifier occurs in the first dynamic conjunct of the sentence. The interpretation assigned to this formula has the same truth conditions as (23b), in which a static (classical) existential quantifier has scope over all the conjuncts.

(23) a. $\exists^{d}x(\text{man}(x) \;\&\; \text{came_in}(x)) \;\&^{d}\; \text{sat_down}(x)$
(for some$^{\text{dynamic}}$x[x is a man and x came in] and$^{\text{dynamic}}$ [x sat down])

b. $\exists x(man(x) \ \& \ came_in\ (x) \ \& \ sat_down(x))$
(for some x[x is a man and x came in and x sat down])

The dynamic implication $\rightarrow^d$ holds between two sentences A and B for a given set R of discourse referents iff every information state which A produces for R gives rise to one which successfully interprets B. The connective $\rightarrow^d$ can be combined with the dynamic existential quantifiers $\exists^d x$, $\exists^d y$ to represent (22b) as (24a), where the dynamic quantifiers occur in the antecedent of the conditional sentence and dynamically bind the variables in the consequent. (24a) has the same truth conditions as (24b), in which the entire conditional is within the scope of two static universal quantifiers.

(24) a. $\exists^d x \exists^d y(man(x) \ \& \ donkey(y) \ \& \ owns(x,y)) \rightarrow^d beats(x,y)$
(for some^{dynamic}x and for some^{dynamic}y[if^{dynamic} x is a man and y is a donkey and x owns y], then^{dynamic} [x beats y])

b. $\forall x \forall y((man(x) \ \& \ donkey(y) \ \& \ owns(x,y)) \rightarrow beats(x,y))$
(for every x and for every y[[if x is a man and y is a donkey and x owns y], then [x beats y]])

(24a, b) are true iff for every pair <a,b> such that a is a man, b is a donkey, and a owns b, a beats b. However, 24a corresponds directly to (22b) in that it represents both indefinite NPs in (22b), *a man* and *a donkey*, as (dynamic) existentially quantified NPs rather than as universally quantified NPs as in (24b). Therefore, this analysis provides an explanation for the fact that, in sentences like (22b), pronouns which are anaphorically dependent upon indefinites behave like variables bound by universal quantifiers.

By defining a dynamic universal quantifier $\forall^d x$ and combining it with $\exists^d y$ and $\rightarrow^d$, it is possible to obtain (25) for (22a). (25) has the same truth conditions as (24a, b). In this formula, *every man* corresponds to a restricted dynamic universal quantifier and *a donkey* to a restricted dynamic existential quantifier.

(25) $\forall^d x((man(x) \ \& \ \exists^d y(donkey(y) \ \& \ owns(x,y))) \rightarrow^d beats(x,y))$ (for every^{dynamic}x [[if ^{dynamic} x is a man and for some^{dynamic}y[y is a donkey and x owns y]], then^{dynamic} [x beats y]])

As in the case of (24a) and (22b), (25) corresponds directly to (22a) in that the indefinite *a donkey* is represented by a (dynamic) existential quantifier rather than a universal quantifier (as in (24b)). Therefore, the dynamic binding account of donkey anaphora also permits us to account for the fact the pronoun *it* in (22a) is understood as bound by a (classical) universal rather than a (classical) existential quantifier.

While DRT uses indefinites to introduce referents into a discourse and dynamic binding relies on dynamic operators to pass information concerning discourse referents from one sentence to another, the third approach locates the mechanism for dynamic anaphora in the interpretation of the pronoun which takes a quantified NP as its antecedent. This sort of pronoun, referred to as an E-type pronoun, effectively functions like a pointer to a description that refers back to an entity (or collection of entities) in the set that is determined by its quantified NP antecedent.[17] Taking *he* in (20) and (22b), and *it* in (22a) and (22b) as E-type pronouns gives interpretations of these sentences corresponding to (26a) and (26b).

(26) a. A man came in. The man who came in sat down.
b. Every man who owns a donkey beats the donkeys he owns.
c. If a man owns a donkey, the man who owns a donkey beats the donkey he owns.

So in (20), for example, the E-type pronoun *he* is interpreted by the description *the man who came in*, as in (26a).

Another way of understanding an E-type pronoun is to treat it as corresponding to a function which applies to objects in an appropriately specified domain to give values in a set defined in terms of the denotation of its antecedent NP. The antecedent of *he* in (20) and (22b) is *a man*, which is not within the scope of a quantified NP. Therefore, the E-type function associated with *he* maps any object in the domain of discourse onto an element in the set of men who own donkeys. In (22a, b) the antecedent of *it* is *a donkey*, which is in the scope of *every man* and *a man*, respectively. It denotes an E-type function from men who own donkeys to (collections of) the donkeys which they own.

The three approaches discussed here use different formal techniques for modeling dynamic anaphora. However, common to all of them is the view that a major part of understanding the meaning of a sentence is knowing its possible influence on the informational structure of a discourse in which it appears.

4 Meanings and Situations: Beyond Possible Worlds

In Section 2, I described the intension of an expression as a function from a possible world to the extension of the expression. A world is the result of assigning the objects of a domain to properties and relations in such a way as to produce a complete state of affairs containing these objects. There are at least some cases where it seems to be necessary to use situations rather than worlds to specify the interpretation of a sentence.[18] A situation is a smaller and more fine-grained object than an entire world. It can be contained in larger situations, and it is, in effect, the specification of part of a world (equivalently, a partial specification of a world).

To see the role of situations in representing meaning let's return to the analysis of generalized quantifiers. In section 2 we characterized the denotation of an NP as a GQ (a set of sets). For quantified NPs, we can, equivalently, take the determiner of the NP as denoting a relation between the set denoted by the noun to which the determiner applies and the predicate set of the VP. For example, the GQ corresponding to *every student* is the set of sets each of which contains the set of students. Alternatively, *every* denotes the relation that holds between any two sets A and B when A is contained in B. On both conditions, (14a) is true iff the set of students is a subset of the set of singers.

(14) a. Every student sings.

Similarly, the determiner *the* denotes the relation that holds for two sets A and B when the unique element of A is a member of B. Therefore, (27) is true iff there is a single woman and she dances.

(27) The woman dances.

If the intension of *the woman* takes the actual world (or any world which resembles it) as its argument, then it will yield the set containing the empty set as the extension of the NP. This is because it is not the case that the set of women has only one element in the actual world. As the relation denoted by *the* does not hold between the set of women and the set of dancers, (27) is false in the actual world. It will only be true in a world containing a unique woman. But this is the wrong result. There are surely cases where an assertion of (27) is literally true in the actual world by virtue of the fact that the speaker is referring to a particular woman, despite the existence of other women in the world.

Instead of treating a property as applying to an object in a world, we can localize the relation to a situation within a world. This will give us statements of the form *Mary is a woman in s*. We can express this relation between a statement and a situation s by saying that s supports the information that Mary is a woman. If we identify a situation s containing a unique woman u and interpret *woman*, relative to s, as denoting the singleton set containing u, then (27) is true if u

dances, even though u is not the only woman in the world. This s is the resource situation which we use to determine the GQ that *the woman* denotes.

Imagine a conversation in which I am telling you about two successive visits to the theater. On the first trip I saw a production of a musical with one female actor, and on the second I saw a comedy which also featured one female actor. Using each play as a resource situation I assert (28a) in describing the first production, and (28b) in my account of the second.

(28) a. The woman sang.
b. The woman did not sing.

Assume, also, that the same actress appears in both plays. It is still the case that both (28a) and (28b) are true. Although the two resource situations identify the same person, each situation supports one of the assertions.

Cooper (1996) uses resource situations to characterize the class of GQs denoted by NPs in natural language. He also points out that it is necessary to distinguish between the resource situation in which the restriction (common noun) set of a GQ is fixed and the situation in which the entire sentence containing the GQ expression is evaluated. (29) brings out the distinction clearly.

(29) Everyone spoke to John.

The quantificational situation q which supports (29) includes John. Therefore, if we identify it with the resource situation r for setting the restriction set of the relation denoted by *every*, (29) implies that John spoke to himself. This consequence is avoided if r and q are distinct. We could, for example, take r to be properly contained in q, so that the restriction set is a subset of the set of people in q.

Cooper also argues that the quantificational situation must be distinguished from the individual situations i in which the property expressed by the VP applies to each of the elements of the restriction set. He invokes cases in which perception verbs, like *see*, take quantified complements to motivate this claim.

(30) a. John saw everyone leave the concert.
b. John saw each person leave the concert.

(30a) can be true in a situation in which John saw all of the people at a concert leave the hall, but he did not observe each person leave individually. By contrast, (30b) is true only if he saw each person leave. This difference in interpretation consists in the fact that the truth conditions for (30b) require the identification of q and i while those for (30a) do not.

Conditional donkey sentences in the scope of quantificational adverbs like *usually* provide another case in which situations play a central role in determining the meaning of quantifier terms. (31) allows at least two different interpretations.

(31) Usually if a man owns a donkey, he beats it.

On one reading, (31) says that for most pairs <a,b> where a is a man, b is a donkey, and a owns b, a beats b. Given this interpretation, (31) is true in the following state of affairs. There are 10 donkey owners, 9 of whom each owns a single donkey, and one who owns 20. The 9 men who each own a donkey do not beat it, but the one donkey owner who has 20 beats all of them. There are 29 distinct pairs of men and donkeys they own. The man who owns 20 is the first element of 20 pairs, with each of his donkeys as the second element of one of these pairs. The 9 other owners and their donkeys contribute the remaining 9 pairs. The sentence is true because the first element beats the second in 20 out of 29 of these pairs. On the second reading, 31 claims that most men who own donkeys beat the donkeys they own. With this interpretation, the sentence is false in the situation described here, as it requires there to be more than 5 men who beat the donkeys they own.[19]

It is possible to account for these interpretations by treating adverbs like *usually* as quantifiers that denote relations between sets of situations.[20] *Usually* denotes a relation that holds between two sets of situations A and B iff most of the elements of A are also in B. The different readings are generated by varying the size of the situations in the restriction set that corresponds to the antecedent of the conditional sentence. If this set contains only minimal situations involving a donkey owner and a single donkey, then (31) asserts that most situations consisting of a man and a single donkey which he owns are situations in which he beats that donkey. This yields the first reading. When the restriction set contains maximal situations involving a man and all of the donkeys he owns, then (31) states that most situations in which a man owns donkeys are situations in which he beats the donkeys he owns. This provides the second reading. The first interpretation is symmetrical in that *usually* quantifies over situations defined by pairs of donkey owners and individual donkeys. The second is asymmetrical as *usually* effectively quantifies only over donkey owners.

By using situations to specify the extensions of predicates and quantificational expressions it is possible to represent aspects of interpretation which cannot be captured in classical intensional semantics.

5 Underspecified Representations: Beyond Compositionality

As we observed in Section 2, the condition of compositionality requires that the meaning of any expression P be computable by a function which, given the meanings of P's syntactic constituents as its arguments, yields P's meaning as its value. We have also seen that Montague grammar satisfies this condition by characterizing the relation between the set of syntactic categories and the set of semantic types as a homomorphism which maps each syntactic structure into a single denotational type.[21] In this framework the meaning of an expression is fully determined by (a) its syntactic structure and (b) the meanings of its constituents.

In order to sustain a homomorphism of this kind, the function which specifies the mapping from syntax to semantics must apply to expressions with fully specified syntactic representations and yield unique semantic values. Therefore, syntactic and semantic ambiguity are eliminated by the mapping which the function specifies. Ambiguous lexical items are divided into words which stand in a one-to-one correspondence with the distinct senses of the original term. The verb *run*, for example, becomes a set of verbs each of which is assigned a denotation corresponding to one of *run*'s meanings (move quickly, operate or administer something, flow, function, etc.).

(32) is ambiguous between two scope interpretations of the quantified NP *a painting* relative to the intensional verb *seek*.

(32) John is seeking a painting.

If *a painting* receives narrow scope relative to *seeks*, then John wants there to be some painting or other which he finds. If it has wide scope, then there is a particular painting which he is looking for. Montague generates these readings from distinct syntactic structures. The narrow scope reading is obtained when *a painting* originates in situ as an argument of *seeking*. For the wide scope reading, *a painting* is generated outside of the sentence *John seeks it* and is substituted for the pronoun. The VP of the first structure denotes the set of things which stand in the seek relation to the intension of the GQ denoted by *a painting*. This set is the value that the function denoted by *seeks* assigns to the intension of *a painting*. On the second syntactic derivation, *a painting* is interpreted as a GQ which applies to the predicate set containing the objects that John is seeking. This derivation yields the interpretation that there is a painting x and John is seeking x.

In fact, it is possible to construct a semantic system that is noncompositional, but relates the meaning of an expression systematically and incrementally to the meanings of its parts.[22] This is

achieved by allowing the mapping from syntax to semantics to be a relation which assigns more than one meaning to an expression under a single syntactic representation. In such a system the verb *run* could be paired with a disjunction of meanings corresponding to each of its senses. (33) would be represented as having one syntactic structure, with a VP headed by a single verb *run*, which is associated with at least two distinct semantic representations.

(33) John ran the marathon.

On one, John was a runner in a race, and on the other he administered it. On this view, (33) would be represented by a single syntactic structure which is mapped to a set containing two interpretations, each providing a distinct set of truth conditions. To obtain a disambiguated reading of the sentence it is necessary to select one element of the set.

A more interesting case of noncompositional interpretation involves mapping a syntactic structure into a set of alternative scope readings. There are at least two ways of doing this. On one approach, quantified NPs can either be taken as GQs in situ (in the argument positions where they appear in the syntactic structure of the sentence) or interpreted through the device of quantifier storage.[23] When storage applies to an NP, a variable meaning is substituted for the argument position which it occupies, and the GQ is placed in a stored part of the meaning of the expression where the NP appears. The nonstored meaning of the expression, which includes the variable in the original argument position of the NP, is combined with the meanings of larger expressions until a point is reached where a predicate set is specified. The GQ can be released from storage at this point and applied to the predicate. As we have seen, if *a painting* in (32) is interpreted in situ, it is within the scope of the verb *seeks* and the narrow scope reading results. If it is placed in storage, the set of objects x such that John seeks x is computed as the interpretation of the open sentence (predicate) *John is seeking x*. The GQ denoted by *a painting* is released from storage and applied to this set to yield the wide scope reading of the sentence. Unlike Montague's analysis, this account assigns a single syntactic structure to (32) where *a painting* is always in object position. The structure is associated with two distinct scope interpretations obtained by different procedures.

On the second approach, sentences containing scope-taking expressions are assigned schematic semantic representations in which the scopes of these terms are left unspecified.[24] In the representation for (32), for example, the scope relation between *a painting* and *is seeking* is undefined. Similarly, *a student* and *every program* are unordered for relative scope in the representation assigned to (34).

(34) A student checked every program.

The second treatment of scope ambiguity is similar to the first in that it also associates a syntactic structure with a set of alternative scope interpretations. However, it implies a more far reaching revision of the compositional view of semantic interpretation. This approach takes the meaning of an expression to be a partial representation R defined in terms of a minimal set of conditions C on the interpretation of R. To obtain a more specified meaning one adds new constraints to C to restrict the set of interpretations with which R is compatible. A compositional semantics provides a homomorphism for mapping unambiguous syntactic structures into fully specified semantic values. An underspecified semantics, by contrast, establishes a relation between syntactic structures and partial semantic representations whose parameters characterize sets of possible values. These sets can be further restricted by adding constraints to the representation.

6 Conclusion

Initial attempts to construct a formal semantic theory for natural language used the syntax and truth definitions of first-order languages as a model. Therefore, they associate the categories of

natural language with the semantic types of first-order logic. Montague introduced a richer type system which permits a direct mapping of complex functional categories into corresponding types. It also expresses the interpretation of higher-order expressions, such as nonfirst-order generalized quantifiers. Dynamic semantics then moved beyond the static meaning of an individual sentence taken in isolation to representing semantic content in terms of the way in which a sentence transforms the information state inherited from previous sentences in a discourse. Situation semantics replaced the interpretation of expressions relative to a possible world with evaluation in a situation, where the latter is a more finely structured and partially specified entity than the former. Finally, underspecified semantics discards the condition of compositionality to construct a more flexible mapping between syntactic structure and semantic interpretation. This approach sustains a systematic connection between the meaning of a phrase and the meanings of its parts while using partially defined representations to capture ambiguity and underdetermined interpretation.

It is important to recognize that as new paradigms of semantic representation have emerged, the leading ideas of the earlier programs have not disappeared. They have continued to survive in various formulations and to exert influence on successive generations of theorists, many of whom attempt to solve semantic problems by integrating the insights of earlier models into new frameworks.

In considering the recent history of semantic theory, it becomes clear that recent years have seen considerable progress in the application of increasingly sophisticated formal techniques to the explanation of a wide range of semantic phenomena. This work has opened up new areas of investigation and yielded promising results which have turned formal semantics into a well-grounded and exciting domain of linguistic research.

NOTES

1 There are numerous introductory texts on formal semantics, each highlighting different issues and tending to represent a particular theoretical paradigm. Chierchia and McConnell-Ginnet (1990), and Heim and Kratzer (1998) are two recent texts which offer interesting background and perspectives on the field. The papers in Lappin (1996a) provide introductions to current research in the major areas for formal semantics.

2 Rules are recursive if they can apply to their own output an unlimited number of times. By virtue of this property recursive rules can generate an infinite number of structures.

3 For a discussion of the relation between the semantics of declarative and nondeclarative sentences see Lappin (1982). For analyses of the semantics of interrogatives see Karttunen (1977), Hamblin (1973), Ginzburg (1996), Higginbotham (1996), and Groenendijk and Stokhof (1997).

4 See Frege (1879), (1891), and (1892). A logic is first-order when all of its predicates (property terms) apply only to individuals in the domain of discourse (the domain of the logic). A higher-order logic contains predicates which apply to properties or sets of individuals (and possibly other higher order entities). So, for example, "green" is a first-order predicate that applies to physical objects, while "partially ordered" is a higher-order predicate of sets.

5 See note 4 for the notion of first-order terms and first-order logic. A first-order language is a formal language all of whose predicates are first-order. We can say that a logic is a formal language which has additional principles that identify a set of sentences in that logic as true.

6 See Davidson (1967a) and the papers in Davidson (1984). For applications of Davidson's program within linguistic semantics see Higginbotham (1985), May (1991), and Larson and Segal (1995). Sher (1991) and (1996) extends Tarskian semantics beyond first-order systems.

7 Interestingly, Tarski expressed skepticism about the prospects for developing formal truth definitions for natural languages. He claimed that their terms are often vague or ambiguous. Moreover, they permit self-reference in a way which generates paradox, as in the famous liar paradox *This statement is false*, understood as referring to itself. Davidson, like most semanticists, attempts

to get around these reservations by adopting an incremental program on which a formal truth definition is first constructed for a representative fragment of a natural language and then extended to progressively larger sets of sentence types.

8 Dowty, Wall, and Peters (1981) provides a very clear and detailed introduction to Montague grammar.

9 For recent introductions to categorial grammar see Moortgat (1988), Morrill (1994), and Jacobson (1996).

10 See Davidson (1967b) for this analysis of adverbs. Higginbotham (1985) proposes a Davidsonian treatment of modifiers within the framework of Chomsky's (1981) government and binding model of syntax.

11 For discussions of generalized quantifiers in natural language see Barwise and Cooper (1981), Keenan and Moss (1984), Keenan and Stavi (1986), van Benthem (1986), Westerståhl (1989), Keenan (1996), and Keenan and Westerståhl (1997). For a comparison of the Davidsonian and the GQ approaches to the semantics of NPs see Lappin (1996b) and (1998).

12 I am again simplifying the account by taking transitive verbs to denote functions on the extensions rather than the intensions of NPs. See Cooper (1983) for a treatment of transitive verbs as functions of this kind.

13 See Barwise and Cooper (1981) and Keenan (1996) for this result.

14 Geach (1962) introduced these sorts of cases into the modern semantics literature. The pronouns which are dependent upon indefinite NPs in (22) are generally referred to as *donkey pronouns*, and the anaphoric relation in these structures is described as *donkey anaphora*.

15 DRT was first proposed by Kamp (1981). An alternative version of this theory is presented in Heim (1982). For a recent model of DRT see Kamp and Reyle (1993).

16 The version of dynamic binding which I am summarizing here is essentially the one presented in Groenendijk and Stokhof (1990) and (1991). For an alternative account see Chierchia (1995). Groenendijk, Stokhof, and Veltman (1996) propose a theory of update semantics based on dynamic binding. Kempson, Meyer-Viol, and Gabbay (2000) develop a deductive approach to dynamic semantics which has much in common with all three approaches discussed here.

17 Evans (1980) initially proposed the idea of an E-type pronoun. Cooper (1979), Lappin (1989), Heim (1990), Neale (1990), Chierchia (1992), and Lappin and Francez (1994) suggest different E-type accounts of donkey anaphora.

18 Barwise and Perry (1983) introduced a situation-based theory of meaning into formal semantics. For more recent work in situation semantics see Barwise (1989), Barwise et al. (1991), Cooper et al. (1990), Gawron and Peters (1990), Aczel et al. (1993), Cooper, Mulai, and Perry (1994), and Cooper (1996). The treatment of generalized quantifiers in terms of situation theory discussed here is based on Cooper (1996).

19 Explaining these distinct readings for (31) is known as the *proportion problem* for conditional donkey sentences with quantificational adverbs of nonuniversal force. See Kadmon (1990), Heim (1990), Chierchia (1992), and Lappin and Francez (1994) for discussions of this problem.

20 Heim (1990), and Lappin and Francez (1994) pursue this approach. Lappin and Francez analyze quantificational adverbs as generalized quantifiers on sets of situations.

21 A homomorphism is a functional mapping from a domain A to a range B in which several elements of A can be associated with one object in B. Montague's category–type correspondence is a homomorphism because in some cases the same semantic type is assigned to more than one syntactic category. For example, both common nouns and predicates denote sets of individuals.

22 See Nerbonne (1996) for a noncompositional approach to semantics in a constraint-based framework. My discussion of compositionality in this section owes much to his treatment of the issue. Zadrozny (1994) shows that any mapping from syntax to semantic interpretation for a language can be formulated as a function, and so can be expressed compositionally. However, such functions may be nonsystematic in the way in which they specify the dependence of a phrase's interpretation on the meanings of its constituents. Specifically, they may involve a case by case listing for subsets of the relevant ordered pairs of meanings for which the functional relation holds.

23 See Cooper (1983), Pereira (1990), Pereira and Pollack (1991), and Dalrymple, Shieber, and Pereira (1991) for accounts of quantifier storage. Lappin (1991) and (1996b) gives arguments for using storage rather than a syntactic operation of quantifier raising to capture wide scope readings of quantified NPs.

24 See Reyle (1993) and Copestake, Flickinger, and Sag (1997) for different versions of this view.

REFERENCES

Barwise, J. (1989). *The Situation in Logic*. Stanford, CA: CSLI.

Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4, 159–219.

Barwise, J., Gawron, J. M., Plotkin, G., and Tutiya, S. (eds) (1991). *Situation Theory and Its Applications*, Vol. 2. Stanford, CA: CSLI.

Barwise, J. and Perry, J. (1983). *Situations and Attitudes*. Cambridge, MA: MIT Press.

Carnap, R. (1947). *Meaning and Necessity*. Chicago, IL: University of Chicago Press.

Chierchia, G. (1992). Anaphora and dynamic binding. *Linguistics and Philosophy*, 15, 111–83.

Chierchia, G. (1995). *Dynamics of Meaning*. Chicago, IL: University of Chicago Press.

Chierchia, G. and McConnell-Ginet, S. (1990). *Meaning and Grammar*. Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Cooper, R. (1979). The interpretation of pronouns. In F. Heny and H. Schnelle (eds), *Syntax and Semantics*, Vol. 10 (pp. 61–92). New York: Academic Press.

Cooper, R. (1983). *Quantification and Syntactic Theory*. Dordrecht: Reidel.

Cooper, R. (1996). The role of situations in generalized quantifiers. In Lappin (1996c), pp. 65–86.

Cooper, R., Mulai, K. and Perry, J. (eds) (1990). *Situation Theory and Its Applications*, Vol. 1. Stanford, CA: CSLI.

Copestake, A., Flickinger, D., and Sag, I. (1997). Minimal recursion semantics: an introduction. MS, Stanford University.

Dalrymple, M., Shieber, S., and Pereira, F. (1991). Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14, 399–452.

Davidson, D. (1967a). The logical form of action sentences. In N. Rescher (ed.), *The Logic of Decision and Action* (pp. 81–120). Pittsburgh, PA: University of Pittsburgh Press.

Davidson, D. (1967b). Truth and meaning. *Synthese*, 17, 304–23.

Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

Dowty, D., Wall, R., and Peters, S. (1981). *Introduction to Montague Semantics*. Dordrecht: Reidel.

Evans, G. (1980). Pronouns. *Linguistic Inquiry*, 11, 467–536.

Frege, G. (1879). Begriffsschrift. In Geach and Black (1970), pp. 1–20.

Frege, G. (1891). Function and concept. In Geach and Black (1970), pp. 21–41.

Frege, G. (1892). On sense and reference. In Geach and Black (1970), pp. 56–78.

Geach, P. (1962). *Reference and Generality*. Ithaca, NY: Cornell University Press.

Geach, P. and Black, M. (eds) (1970). *Translations from the Philosophical Writings of Gottlieb Frege*. Oxford: Blackwell.

Ginzburg, J. (1996). Interrogatives: questions, facts, and dialogue. In Lappin (1996c), pp. 285–322.

Groenendijk, J. and Stokhof, M. (1990). Dynamic Montague grammar. In L. Kalman and L. Polos (eds), *Papers from the Second Symposium on Logic and Grammar* (pp. 3–48). Akademiai Kiado, Budapest.

Groenendijk, J. and Stokhof, M. (1991). Dynamic predicate logic. *Linguistics and Philosophy*, 14, 39–100.

Groenendijk, J. and Stokhof, M. (1997). Questions. In Van Benethem and ter Meulen (eds), pp. 1055–124.

Groenendijk, J., Stokhof, M., and Veltman, F. (1996). Anaphora, discourse, and modality. In Lappin (1996c), pp. 179–213.

Hamblin, C. L. (1973). Questions in Montague grammar. *Foundations of Language*, 10, 41–53.

Heim, I. (1982). The semantics of definite and indefinite noun phrases. PhD dissertation, University of Massachusetts, Amherst, MA.

Heim, I. (1990). E-type pronouns and donkey anaphora. *Linguistics and Philosophy*, 13, 137–77.

Heim, I. and Kratzer, A. (1998). *Semantics in Generative Grammar*. Oxford: Blackwell.

Higginbotham, J. (1985). On semantics. *Linguistic Inquiry*, 16, 547–94.

Higginbotham, J. (1996). The semantics of questions. In Lappin (1996c), pp. 361–83.

Jacobson, P. (1996). The syntax–semantics interface in categorial grammar. In Lappin (1996c), pp. 89–116.

Kadmon, N. (1990). Uniqueness. *Linguistics and Philosophy*, 13, 273–324.

Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, and M. Stokhof (eds), *Formal Methods in the Study of Language* (pp. 277– 322). Amsterdam: Mathematical Center Tracts.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. Dordrecht: Kluwer.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, 1, 3–44.

Kasher, A. (ed.) (1991). *The Chomskyan Turn*. Oxford: Blackwell.

Keenan, E. (1996). The semantics of determiners. In Lappin (1996c), pp. 41–63.

Keenan, E. and Moss, L. (1984). Generalized quantifiers and the expressive power of natural language. In J. van Benthem and A. ter Meulen (eds), *Generalized Quantifiers* (pp. 73–124). Dordrecht: Foris.

Keenan, E. and Stavi, J. (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy*, 9, 253–326.

Keenan, E. and Westerstahl, D. (1997). Generalized quantifiers in linguistics and logic. In van Benthem and ter Meulen (1997), pp. 838–93.

Kempson, R., Meyer-Viol, W., and Gabbay, D. (2000). *Dynamic Syntax: The Deductive Flow of Natural Language*. Oxford: Blackwell.

Lappin, S. (1982). On the pragmatics of mood. *Linguistics and Philosophy*, 4, 559–78.

Lappin, S. (1989). Donkey pronouns unbound. *Theoretical Linguistics*, 15, 263–86.

Lappin, S. (1991). Concepts of logical form in linguistics and philosophy. In Kasher (1991), pp. 300–33.

Lappin, S. (1996a). The interpretation of ellipsis. In Lappin (1996c), pp. 145–76.

Lappin, S. (1996b). Generalized quantifiers, exception phrases, and logicality. *Journal of Semantics*, 13, 197–220.

Lappin, S. (ed.) (1996c). *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell.

Lappin, S. (1998). Semantic types for natural language. Inaugural lecture, SOAS, University of London.

Lappin, S. and Francez, N. (1994). E-type pronouns, I-sums, and donkey anaphora. *Linguistics and Philosophy*, 17, 391– 428.

Larson, R. and Segal, G. (1995). *Knowledge of Meaning*. Cambridge, MA: MIT Press.

May, R. (1991). Syntax, semantics, and logical form. In Kasher (1991), pp. 334–59.

Montague, R. (1974). The proper treatment of quantification in ordinary English. In R. Montague, *Formal Philosophy*, ed. R. Thomason (pp. 247–70). New Haven, CT: Yale University Press.

Moortgat, M. (1988). *Categorial Investigations*. Dordrecht: Foris.

Morrill, G. (1994). *Type-logical Grammar*. Dordrecht: Kluwer.

Neale, S. (1990). *Descriptions*. Cambridge, MA: MIT Press.

Nerbonne, J. (1996). Computational semantics. In Lappin (1996c), pp. 461–84.

Pereira, F. (1990). Categorial semantics and scoping. *Computational Linguistics*, 16, 1–10.

Pereira, F. and Pollack, M. (1991). Incremental interpretation. *Artificial Intelligence*, 50, 37–82.

Reyle, U. (1993). Dealing with ambiguities by underspecification: construction, representation and deduction. *Journal of Semantics*, 10, 123–79.

Sher, G. (1991). *The Bounds of Logic*. Cambridge, MA: MIT Press.

Sher, G. (1996). Semantics and logic. In Lappin (1996c), pp. 511–37.

Tarski, A. (1933). The concept of truth in formalized languages. In A. Tarski (1983), *Logic, Semantics and Mathematics*, 2nd edn (pp. 152–278), translated by J. H. Woodger, edited by J. Corcoran. Indianapolis, IN: Hackett.

van Benthem, J. (1986). *Essays in Logical Semantics*. Dordrecht: Reidel.

van Benthem, J. and ter Meulen, A. (eds) (1997). *Handbook of Logic and Language*. Amsterdam: Elsevier.

Westerståhl, D. (1989). Quantifiers in formal and natural languages. In D. Gabbay and F. Guenther (eds), *Handbook of Philosophical Logic*, Vol. 4 (pp. 1–131). Dordrecht: Reidel.

Zadrozny, W. (1994). From compositional to systematic semantics. *Linguistics and Philosophy*, 17, 329–42.

15 Historical Linguistics

Language Change Over Time

BRIAN D. JOSEPH

1 Introduction

One remarkably striking observation about language, seemingly trivial but actually quite important, is that languages change through time. It is at least conceivable that language could remain unchanged over time, as is the case with some other human institutions, e.g., various taboos or the rules to some games, and with some aspects of human communication systems, e.g., Morse Code or the value of a smile as a nonverbal signal,[1] but the facts tell us otherwise.

The mutability of languages can be demonstrated empirically through a comparison of a single language at different stages in its history. For instance, (1) below provides first lines of some great works from three periods of English: Old English as represented by Caedmon's hymn of the seventh century CE, Middle English as represented by Chaucer's Prologue to the *Canterbury Tales* from the late fourteenth century, and early Modern English as represented by Shakespeare's *Othello* from the early seventeenth century:

(1) English at various stages in its history

a. **Nū wē sculon herian heofon-rīces Weard** … (Caedmon, *Hymn*, ca. 660)
'Now we ought to praise the guardian of the kingdom of heaven'

b. **Whan that Aprille** with **his shoures soote** … (Chaucer, *Canterbury Tales*, ca. 1400)
'When April with its sweet showers …'

c. Tush, never tell me! I take it much unkindly that **thou**, Iago,
who **hast** had my purse as if the strings were **thine, shouldst**
know of this. (Shakespeare, *Othello*, 1604)
'Bah, never tell me! I take it much unkindly that you, Iago, who has had my purse as if the strings were yours, should know of this.'

The boldface in (1) marks those features – pronunciations (as reflected in the spelling), words, sentence and phrasal constructions, and the like – which are not part of contemporary English usage. As the translations show, the differences are considerable and noticeable. For instance, the long monophthongal vowels of *nū* and *wē* in (1a) – assuming that such is the correct interpretation

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

of the spelling – are pronounced as diphthongs in their modern counterparts *now* and *we*, respectively; *sculon* in (1a) shows a plural form absent in its modern counterpart *shall*; *whan that* in (1b) has two subordinating elements (a doubly filled COMP(lementizer) node, in some interpretations) where the modern counterpart *when* has only one; and forms such as *tush*, *thou*, and *hast* of (1c), while marginally possible in present-day English, are certainly not at all usual. Significantly, examples like these, reflecting change in the language over a period of some 1,300 years, can be found in language after language for which records prior to the contemporary period exist; nor must the time depth be great to reflect change – comparing Mark Twain's nineteenth-century usage *I am become* with twentieth-century *I have become* reveals a change in the selection of auxiliary verbs in the perfect tense of *become* within a span of approximately a hundred years, and the current use of *be like* to introduce direct speech (e.g., *And I'm like "Oh my God!"*) seems to have arisen since the 1970s,[2] and is replacing the earlier colloquial use of *go* (e.g., *And I go "Oh my God!"*).

Moreover, it does not take a trained specialist to be aware of language change. Over the years, again and again, similar observations been made by nonlinguists, offering further support for recognizing the ubiquity of change in language. For instance, Socrates, as reported by Plato in the *Cratylus* (418C), commented on what he (incorrectly) analyzes as a conservative pronunciation on the part of women of his day compared to the pronunciation of others, which he mistakenly saw as innovative:[3]

> You know that our ancestors made good use of the sounds of iota [a vowel letter of the Greek alphabet] and delta [a consonant letter], and that is especially true of the women, who are most addicted to preserving old forms of speech. But nowadays people change iota to eta or epsilon [two other vowels], and delta to zeta [another consonant], thinking they have a grander sound. … For instance, in the earliest times they called day *himéra*, others said *heméra*, and now they say *hēméra*.

As Teodorsson (1979: 69) notes, all the evidence known now indicates that *hēméra* is the older pronunciation of 'day' in Ancient Greek, so the proper interpretation of Socrates' observations is that "the i-pronunciation used by women was that of the innovative phonological system" and thus that this innovative pronunciation coexisted as part of a change in progress with the more conservative *heméra* and *hēméra*.

And Chaucer himself remarked on the language of a thousand years before him in a famous passage from *Troilus and Creside* (II.22–28):[4]

> Ye knowe ek that in forme of speche is chaunge
> Withinne a thousand yeer, and wordes tho
> That hadden pris, now wonder nyce and straunge
> Us thinketh hem, and yet thei spake hem so,
> And spedde as wel in love as men now do;
> Ek for to wynnen love in sondry ages,
> In sondry londes, sondry ben usages.
>
> 'You know also that there is change in the form of speech within a thousand years, and of words at one time, that had value, now wondrous foolish and strange to us they seem, and yet they spoke them thus, and they prospered as well in love as men now do; also for winning love in various times, in various lands, various were the usages.'

All of these examples thus attest to change being a continuing force in language. Historical linguistics is the branch of linguistics that is concerned with language change in general and with specific changes in languages, and in particular with describing them, with cataloging them, and, ultimately, with explaining them. Thus in addition to looking at language change, historical linguistics is also interested in language history, i.e., in working out the details of how particular languages develop through time. Somewhat paradoxically, a concern for language

history means that change is not the only focus of historical linguistics; in the course of time, while virtually all aspects of a language, excepting those that correspond to truly inviolable linguistic universals, can in principle change, some aspects of a language may remain stable and not change. In fact, for some linguists, unchanging elements in a language may provide important clues regarding its (pre)history (see Section 6).

To return to Socrates' linguistic comments in the *Cratylus*, he was really engaging in the observation of language change in the example cited above, since, under Teodorsson's interpretation, he was attending to variation evident synchronically around him in Greece of the fifth century BCE. Chaucer, on the other hand, in his musings in *Troilus and Creside*, was engaging in an exercise in language history, by speculating on what earlier stages of English had been like. As should be clear, both types of pursuits have their place in historical linguistics. The study of synchronic variation, though associated with quantitative sociolinguistics (see Chapters 24 and 25), is a window into change in progress, especially on the assumption that an innovation, whether internally caused or introduced through contact with speakers of other languages, starts in a restricted part of a speech community and then spreads (see Section 5); on the other hand, the study of language history is a window, perhaps a speculative one, into the past, and it is associated with reconstruction of earlier language states and with working out the relationships among languages that give clues to how they came to be as they are. Moreover, in order to understand the history of particular languages, one has to have some assumptions in place as to how languages can change, for otherwise there is no framework for analyzing observed or hypothesized changes, or the movement from one language state, whether attested or hypothesized (i.e., reconstructed), to another.

These two aspects of historical linguistics are linked also by the so-called "uniformitarian principle," which states (in the formulation of Hock 1991: 630): "The general processes and principles which can be noticed in observable history are applicable in all stages of language history." There may well be reason to believe that the bases for this principle are suspect,[5] in that, for instance, processes of change observable in modern urban settings need not be evident or have been operative in pastoral communities of millennia ago. Still, we do know that humans today and humans 4,000 or so years ago are not all that different physically, to judge from burial remains, and emotionally, to judge from themes in ancient literature, so that some parallelism in regard to language behavior would not be unexpected.[6] Moreover, with this principle, observing change in progress in the present day provides insights that can be used for unraveling aspects of language development in the past into which we often have no other basis for insight; that is, with the "uniformitarian principle," we are licensed to make educated guesses about the past generated by our study of the present.

2 Framing the Issues

To set the stage for the discussion to follow and by way of framing the various issues to be considered, we turn to five key questions concerning language change, the problems which Weinreich, Labov, and Herzog (1968) say that "a theory of change must solve"; as restated and elaborated by Labov (1982), these problems are: the "constraints" problem, the "transition" problem, the "embedding" problem, the "evaluation" problem, and the "actuation" problem.

The "constraints" problem focuses on what general constraints on change, if any, there are that determine possible and impossible changes and directions of change. One side of this problem, as put in the restatement by Labov (1982), focuses on how a solution "would advance our understanding of the causes of change, since each constraint demands an explanation, and that explanation will usually bear on the cause of the change." There is also a purely descriptive side to this question in that knowing the inventory of changes that have occurred is the first step toward understanding what the range of possible changes is and thus what the impossible changes

are. In this way, a third side to the "constraints" problem emerges, for it allows for an important connection to be made between diachronic linguistics, the examination of language through time, and synchronic linguistics, the analysis of a language at any given point in time.

One way of stating the goal of (synchronic) linguistic theory is that it aims to characterize the class of possible human languages, thereby ruling out those linguistic states which never occur and are "impossible" human languages. Moreover, the way most linguists have attempted to achieve that synchronic goal is to identify a set of linguistic universals. Now, in doing synchronic analysis we usually identify a "slice" of a language at a particular point in time, but clearly, the "point" in question is arbitrary and can be cut finely or broadly. Thus, while English of the twenty-first century forms a synchronic "slice" that we can examine, so does Modern English, defined from Shakespeare's time in the late sixteenth century to the present, and so does English of the 1980s, etc. With this view of synchrony, diachrony can be defined as the transition through successive, finely cut synchronic states, and can be schematized as follows:

D	L1	Synchronic Stage 1
I	L2	Synchronic Stage 2
A	L3	Synchronic Stage 3
C	L4	Synchronic Stage 4
H	.	.
R	.	.
O	.	.
N	Ln	Synchronic Stage n
Y	Ln+1	Synchronic Stage n+1

Linguistic universals, assuming they can be determined, hold at each synchronic stage and define "possible" and "impossible" human languages at each stage. Presumably, also, they hold in the transition between synchronic stages, inasmuch as the division between these stages is arbitrary, and diachrony forms a continuum of synchronic stages. Under such a view, therefore, with an appropriate set of universals, the "constraints" problem of determining possible and impossible changes reduces to the synchronic question of determining possible and impossible human languages. In a sense, then, the two pursuits are the same, and this view of the relationship between synchrony and diachrony makes it clear just how similar they are.

The "transition" problem seeks to answer by what route language changes. The interest here is similar to the view in the above diagram, for a "dynamic perspective" is needed to allow for a seamless movement through successive synchronic states. As Labov (1982: 28) notes, in essence, "solutions to the transition problem can be restated as solutions to the problem, 'How can language change from one state to another without interfering with communication among members of the speech community?'"

There is yet another direction in which this question can be taken, i.e., expressing an interest in the specific paths followed by a change: Does a change from X to Z necessarily go through an intermediate stage Y? For example, in the transition from Old English [ēː] (as in *wēː* in (1a)) to Modern English diphthongal [ij] (as in *we*), must there have been an intermediate stage of [iː] or [ej] or the like, or could [ēː] become [ij] directly?

The "embedding" problem focuses on how a given language change is embedded in the surrounding system of linguistic and social relations. This issue on the one hand asks whether there are system-internal conditions that induce or inhibit change. For example, is the packing of several sounds into a relatively small acoustic and articulatory space (as with Serbian voiceless affricates: dental [c], alveopalatal [ć], and palatal [č]) likely to lead to a loss of some of these distinctions?[7] On the other hand, since conditions external to the linguistic system, e.g., social unrest, wars, forced migrations, etc., could also conceivably contribute to or affect change in

language,[8] this issue, together with the evaluation problem, sets the study of language change squarely within the social arena.

The "evaluation" problem asks how members of a speech community evaluate a given change, and what the effect of this evaluation is on the change. Here the focus is preeminently sociolinguistic in nature, for any innovation in a speaker's linguistic usage that is salient and perceptible – whether it is a new turn of phrase or new lexical item, a new pronunciation, a new syntactic construction, a new meaning for an already existing word – can evoke an evaluative response from the hearer: Is this innovation one that I as a speaker like, one that I might now choose to adopt in my own speech, or is it one I would want to avoid? Language use in this view says something about each of us as individuals and as members of a group, and this social dimension to language use turns out to be crucial to understanding language change and especially the spread of innovations.

Finally, there is the "actuation" problem of why a given linguistic change occurred at the particular time and place it did. This problem seeks to find the conditions that lead to a given change, and adds a further dimension to the understanding of language change, for if we understand the causes of change well enough and can pinpoint certain conditions present in a speech community and/or a linguistic system, we ought then to be able to "predict" (in a retrospective way, so that perhaps "post-dict" or "retro-dict" would be more appropriate) the direction of change. "Predict" here does not have its usual sense of hypothesizing about what might happen in the future, and, indeed, scholars of language change, perhaps unnecessarily, generally avoid making even educated guesses about future language states; rather, "predict" here means giving an explanation for why a given element in a language – a sound, a form, a construction, etc. – changed the way it did, rather than in some other possible way. For example, why did Old English [ē] become in later English [ij] rather than *e* or *a* or some other vowel?[9]

Several of these foundational questions are interconnected, as the discussion above makes clear, and lend themselves to the statement of other related issues, such as the relation between synchrony and diachrony mentioned in connection with the "constraints" problem. Moreover, other issues not overtly stated by Weinreich, Labov, and Herzog can be mentioned. Particularly vexing is the determination of "naturalness" in the direction of language change: Can change ever lead to an unnatural state? Are some changes more likely to occur than others? Classification of changes and observation of the range of possible changes are clearly of relevance here, but so too is an understanding of the physiological bases for sound change, the psychological bases for morphological change, and the like.

In the sections that follow, these various facets of historical linguistics are explored as the study of both language change and language history; moreover, in so doing, some of the methods used by historical linguists in their investigations are brought to light.

3 Substance of Change: What Types Occur? How Do They Spread?

It is stated above, almost as an axiom, that virtually all aspects of a language are subject to change, except for those that correspond to absolute linguistic universals that truly cannot be violated. Thus, the simple answer to what can change in a language is "(virtually) everything," though it is not the case that everything in a language at a given point must change – there can be diachronic stability as well as diachronic change. For example, except for the realization of the main accent, from high pitch to greater loudness, the Greek word *ánemos* 'wind' has remained virtually unchanged for at least 2,500 years: in its segmental phonological composition, its morphological form, its syntactic behavior, and its meaning.

This simple answer about what can change makes it difficult to exemplify all types of change in a brief discussion, but an examination of any earlier stage of any language along with a

comparison with a later stage will reveal a certain number of changes. Examples are provided here from just two languages, but a similar exercise involving other languages would yield similar results.

Example (1a) from English of 660 CE, as compared with modern English, reveals changes in phonology, e.g., *nū* → *now*, *wē* → *we*; morphology, e.g., absence of plural marking on the verb *sculon*, which ultimately yielded *should*; and lexicon, e.g., the loss of the word *herian*, and the addition of the word *praise*, which entered the language some six centuries later. The changes in the once free word *ric-* 'realm' straddle the boundary between morphology and the lexicon – it is now restricted to occurrence as a bound element, though possibly still recognizably segmentable as a morpheme in *bishopric* 'the diocese or office of a bishop' (segmentable due to the independent existence of *bishop*), but has no clearly recognizable morphemic status in *eldritch* 'strange or unearthly'. Moreover, Chaucer's subordinate clause with *whan that* as opposed to standard Modern English *when* by itself gives an example of a change in sentence structure (syntax).

Similarly, between Ancient Greek and Modern Greek, with regard to phonology and morphology, one finds changes in the realization of sounds, so that [ü, ü:, ē, oi, i, ī] all merged eventually to [i], aspirated voiceless stops [p^h t^h k^h] became voiceless fricatives [f θ x], etc.; and in the form of grammatical endings, e.g., second person past tense imperfective aspect nonactive voice *-so* became *-sun*, matching the first person ending *-mun* in vocalism and final segment. Changes are also evident in the extent of word formation processes, e.g., coordinative compounds of the type *maxero-píruna* 'knife and fork; cutlery' were rare in Ancient Greek but have become more numerous in Modern Greek and the type has been extended to verbs, as in *aniɣo-klíno* 'I open and close'. Further, Greek syntax has shifted drastically, as the infinitive of Ancient Greek has given way to finite-clause replacements, and constructions which once tolerated missing (understood) objects have yielded to ones with overt expression of the object, both illustrated in (2), among other changes:[10]

(2) a. ēn ho trugētos hetoimos tou therizein (*1 Samuel* 13.21 [2nd c. BCE])
was/3SG the-harvest ready COMP harvest/INF
'The harvest was ready for harvesting'
(NB: the object of the infinitive *therizein* is not overtly expressed)

b. ekhei hetoimon ton daon na ton eparei (*Lybistros* 2663 [14th c. CE])
has/3SG ready the-torch COMP it/ACC take/3SG
'She has the torch ready for him to take'
(NB: literally, "… ready that he take it," with a finite complement and the object of *eparei* overtly expressed (*ton*))

Moreover, in keeping with the program suggested above whereby one can learn about language change from synchronic variation, an example from contemporary American English can be cited. In Central Ohio, among younger speakers in the 1960s, the verb *bean* was used in baseball parlance to refer to being hit by a pitched ball on one's head, whereas for younger speakers 30 years later in the 1990s, it refers to being hit with a pitch anywhere on the body, thus with a broader meaning. The synchronic variation in the 1990s between younger speakers with the innovative broad meaning and (now) older speakers with the narrower meaning suggests a change that may ultimately spread across all age groups in the speech community as the now younger speakers age.

From the point of view of the "evaluation" question discussed above, when these innovations, or any innovation, first entered the language, they must have provoked a certain reaction from those who heard them, perhaps even a negative one. Most readers will have had the experience of hearing some technology-oriented neologism for the first time, e.g., *access* as a verb (e.g., *You can access that information electronically*), *e-mail* as a count noun (e.g., *I received 30 e-mails this morning*),

or *e-* as a prefix referring to electronic transmission (as in *e-mail*, *e-trade*, *e-commerce*, *e-talk*, etc.), of needing to decide whether to adopt such usages, and of finding that, even if one winced on first hearing them, repeated use by others made it easy finally just to go along and join in the innovative usage.

In a similar way, though surely with more complicated motivation on the part of adopting speakers, all innovations that ultimately are generalized over the (relevant) speech community must be positively evaluated by speakers and actively (though not necessarily consciously) adopted by them. Such innovations, once they have spread, can be called "real" changes, in that the behavior of the speech community at large has been affected. Significantly, as a corollary, it must be noted that not all innovations take hold and spread so as to become changes in a whole speech community; restricted spread of an innovation can lead to the formation of dialects within a larger speech community. Moreover, not all synchronic variation will result in a change in the long term, for there can be situations in which stable variation persists over long periods of time; for instance, the variable deletion of the past-tense marker *-t/-(e)d* (e.g., *kep'* for *kept*) in American English has been stable for several generations (Labov 1989). The dynamics of the spread of innovations and the resolution of competition between innovative and older variants largely constitute a sociological matter, but, clearly, one with linguistic consequences (see also the end of Section 4).

It is suggested above that at the simplest level, the mere repetition and recurrence of some innovative usages can inure a speaker to their novel nature and thus promote acceptance and eventual adoption and spread. Another dimension to the matter of recurrence of innovations is the fact that some changes are found to occur again and again, independently, in language after language, thus giving a basis for deeming such a change to be a natural one. Some examples of such recurring types of changes include the following:

(3)
a. the change of [f] to [h] occurred in the ancient Italic language of Faliscan, in Spanish, and in some varieties of Chinese (and no doubt elsewhere)
b. devoicing of word-final voiced stops occurred in Russian, Turkish, and German (e.g., earlier *rad* 'wheel' has come to be pronounced [rat])
c. reductions of clusters with concomitant lengthening of an adjacent vowel ("compensatory lengthening"), as in Late Latin *asnu* 'ass' → French *âne* (pronounced [a:n] at first), or Old English *thegn* → Modern English *thane*
d. loss of unaccented vowels, especially word-medially (syncope), as in Middle English trisyllabic *chimenee* (accent on the initial syllable) becoming Modern English disyllabic *chimney*, with similar changes in Latin and Old Irish
e. adjacent sounds coming to agree in certain features (assimilation), as in Old English *hænep* yielding (ultimately) Modern English *hemp*, with the nasal and stop consonants, adjacent after syncope of the unaccented *-e-*, agreeing in point of articulation (both labial, as opposed to dental versus labial earlier); similar changes occur in Greek, Latin, Sanskrit, Arabic, and virtually every other language known
f. reanalysis of third person verb forms with a person-marking suffix as having no suffix (thus as base forms) occurred in Greek, Persian, and Sanskrit[11]
g. in many languages, analogically innovated forms have taken over the primary function for a sign while the forms they replace, if they survive at all, take on a restricted function, as with English *brothers* ousting the older but now functionally quite limited *brethren*, among other cases[12]
h. in many languages, words that were once free and independent have come to take on the status of bound affixes, as in Latin *mente*, the ablative case of 'mind' coming to be the French adverbial suffix *-ment*, as in *clairement* 'clearly' (and thus etymologically, "with a clear mind")[13]

i. the broadening of referent seen in the above example of Central Ohio *bean* recurs in the development of Middle English *dogge* 'hunting dog → Modern English *dog*, referring to canines in general.

Identifying such changes as "natural," and thus unsurprising when they occur is in keeping with Labov's "constraints" problem and the "actuation" problem, as discussed above.

Searching for parallels and deriving inferences about naturalness of developments is thus an important part of historical linguistics, but one has to be cautious about not going too far, in that "natural" need not mean "necessary" or "only in one direction." Thus many languages, including English, persist in having word-final voiced stops quite happily, and some have even undergone word-final voicing, as the evidence of the third person singular past ending *-d* in Old Latin, from Proto-Indo-European *-t shows, and cases of movement from bound affix to independent word (the reverse of the *mente* example) are known.[14] Moreover, in some domains, for instance, semantic change, the directions of changes are so tied to the real-world sociocultural context that being able to label recurring results of changes, as with the cases of broadening mentioned above, does little to actually advance our understanding of why a change occurred. For instance, English *bead* changed in meaning from 'prayer' to 'small round glass object'; such an innovation in the referent associated with a particular form can make sense only in the context of the counting of prayers on rosaries, and so is one that no theory of semantic change could predict as "natural."[15]

4 Mechanisms of Change: How Is Change Manifested in Language?

One way that language change is manifested, clearly, is through changes in the behavior of speakers, in that a word comes to be pronounced in a different way, used in a novel construction, extended in meaning, and so on. In such ways, language change is manifested as alterations in the actual form that language takes in the mouths (or hands)[16] of its users, what might in the terminology of recent decades be termed changes in the surface structure, i.e., in the output of the grammar.

However, for the most part, explicit synchronic accounts of a linguistic phenomenon are necessarily couched in a particular theoretical framework and the formalism associated with that framework. This enterprise is driven by the assumption made by (most) linguists that there is some correct linguistic theory that is operative – we may not yet have found the very best theory, but the exercise of positing analyses and testing them is part of the process that will lead ultimately to the discovery of that best theory. Moreover, given that, as the diagram in Section 2 above indicates, diachrony is the progression through successive synchronic states, and further that the current conception of the "best" linguistic theory is the medium for describing and analyzing the grammars of each of those synchronic language states, it is natural to think that language change can be accounted for or at least best characterized in terms of change in these grammars.

Indeed, in the past 40 years or so, there have been several attempts at devising an account of language change in just those terms: Halle (1962), for instance, equated sound change with the addition of a phonological rule to the end of a grammar; Kiparsky (1968) utilized changes in the form of phonological rules as well as in their order relative to one another as a means of accounting for phonological change; and Klima (1964, 1965) took a similar approach to syntactic change.[17] More recently, with a change in the dominant theoretical paradigm, in phonology in the United States at least, to Optimality Theory, a constraint-based approach to grammar, the view has been advanced that phonological change is the result of changes in the strengths of constraints relative to one another.[18]

This view treats (surface) language change as a function of changes in grammars, and thus a secondary side effect, a derivative, of changes motivated by abstract properties of grammars; still, it is an attractive view, one that is easy to believe in. However, there is good reason to reject it as the right way to view the process and mechanisms of language change; at best, it would seem to provide a convenient description of the difference between two stages of a language. For one thing, as Andersen (1973) has observed, saying that sound change is the addition of a phonological rule to the grammar does not answer the question of where the rule comes from in the first place; he looks instead to the reinterpretation of ambiguous acoustic signals as a possible source of sound change (see also Section 5) and sees rule addition as a construct that describes the diachronic correspondence between the grammar before the reinterpretation and the grammar afterward but does not give any insight into the process(es) that gave rise to the change.

Also, changes in phonological rule systems of the sort that motivated the treatment in Kiparsky (1968) were generally associated with changes in paradigms. For example, an early Latin rule of w → Ø that accounted for the relationship in the root between *par-os* 'little/NOM.SG.MASC' and *parw-ī* 'little/GEN.SG.MASC' is absent from later Latin, and that absence correlated with the appearance of a uniform paradigm in Classical Latin *parw-os/parw-ī* (spelled *paruus/parui*). However, that correlation is a complete accident if the motivation for change resides in abstract properties of a grammar, such as the number of rules a system has,[19] for the loss of a rule would not necessarily lead to a uniform paradigm. On the other hand, as Kiparsky (1971) recognized, one could instead place a positive value on aspects of the output of rules,[20] such as uniformity within a paradigm, and posit that the motivating force for changes in grammars resides in the nature of the output they generate. In that case, the loss of the Latin *w*-deletion rule would be a highly valued event, since the output of the resulting grammar without this rule has a uniform paradigm with *w* in all forms. If that is the case, though, one has to wonder why it is necessary to talk in terms of changes in rules and grammars at all! One could instead view the change in surface forms (e.g., *paros* → *parwos*) as the primary change (on the motivation for which, see Section 5) and then view changes in the form of grammars as at best a description of the comparison of the grammar before the change with the grammar afterward.

Looking at change as something that is manifested in and motivated by a rule system makes it hard to account for changes that have a restricted distribution, for the very notion of "rule" implies some generality over large sets of forms. For instance, as Hock (1991: 256) notes, at least some changes in form motivated by a (psychologically based or analogical) association do not lend themselves well to treatment in terms of rule change, since there are no rules at all involved in the change. He cites the example of so-called "contamination," as seen in the change of French *femelle* to *female* as it was borrowed into English, based on a perceived connection with the semantically close word *male*.[21] Similarly, the early Modern Greek weak third person subject pronoun, e.g., masculine singular *tos*, seems to have originated in a construction with the demonstrative *ná* 'here is/are' and spread from there, but only to use with the locative question word *pún* 'where is/are?'; thus while the use of this innovative form has expanded beyond its original locus, it has not done so to any great extent, so that speaking in terms of the extension of a rule here is not particularly insightful.[22]

As another case of a change that starts in a restricted linguistic environment and then spreads on a limited basis, consider the change by which a *-g-* has come to occur in the first person singular present indicative of certain verbs in Spanish, e.g., *salgo* 'I depart'. This *-g-* appears to have originated in a few verbs where it was the result of regular sound changes, and then to have spread to other verbs on a limited basis. Moreover, with verbs that acquired this *-g-*, it spread within the verbal paradigm in a very limited way, into all forms of the present subjunctive (e.g., *salgas* 'you might depart') but nowhere else, not even other forms of the indicative.[23] It is difficult to see how a rule-based account would be explanatory here, since there is no obvious basis for deriving the subjunctive stem from the first person indicative stem; rather the simple occurrence of a stem allomorph somewhere in the overall paradigm seems to have been basis enough for

a spread into other, even distantly related, forms. The frequency of cases such as these – and examples could be multiplied – suggests that this might be the most accurate model of how change occurs and manifests itself in the grammar of a language and in the behavior (output) of speakers, with the widely seen apparently general changes simply representing the endpoint of a series of limited extensions of a change from its point of origination.

Another dimension to the issue of how change in language is manifested has to do with where change starts in a speech community and where it ends up, as suggested in Section 3. Just as a change might start in a restricted part of the grammar, and be generalized from there, as with the Greek and Spanish examples just mentioned, it is also the case that most changes appear to start in a limited subset of the speech community and then spread from there (if they spread at all), largely driven by social factors such as the prestige (overt or covert) of the group originally identified with the innovative pronunciation, form, construction, turn or phrase, or whatever. This model for change was developed by William Labov, based on his observations of centralization of diphthongs in Martha's Vineyard in the early 1960s, and has been amplified upon in numerous studies since then.[24] Such a model for the spread of an innovation raises an important question that is not fully resolved to every linguist's satisfaction: When is a change said to have occurred, at the first point at which an innovation appears in the speech of some individual or only when the innovation has spread somewhat through at least some part of the speech community? Some linguists see the spread as a purely sociological phenomenon and thus concentrate on what permits the emergence of an innovation in the first place (system-internal factors, contact with other speakers, etc. – see Section 5) while others say that individual perturbations in usage are insignificant unless others adopt them, so that "real" change is only at the level of the speech community, or some subset thereof. It needs to be noted as well that limited spread through a speech community is one basis upon which dialects are created, and if a sufficient number of innovations are shared by some subset of speakers to the exclusion of other parts of the speech community, a separate language can well result.[25]

5 Explanation of Change: Why Does It Happen?

The preceding sections have shown that many different kinds of change in language as well as change at all levels are possible. Consequently, it may seem that change is inevitable, and in some sense it is, in that change is no surprise. Nonetheless, linguists tend to treat the lack of change, i.e., linguistic stability from generation to generation, as the unmarked situation, so that change, when it does occur, demands an explanation. It is useful therefore to consider the various factors that induce change, that is, to explore the underlying causation of language change.

There are four main kinds of factors that play a role in inducing language change: psychological (cognitive) factors, physiological factors, systemic factors, and social factors. These all make sense in that they correspond to different aspects of language: language as a psychological "entity" housed (somewhere) in the brains of speakers (hence cognitive, also); language as the production of sounds and signs and forms through the physiology of the human body (e.g., the vocal tract); language as a system with regularities and interacting components; and, finally, language as a social "organism" that exists in the interactions between and among members of social groups. These various causal factors are briefly introduced in what follows.

Several of the examples discussed above can be explained by reference to psychological or cognitive factors. Key among these is analogy, which can be described as the influence of one form or class of forms over another and is psychological/cognitive in that it really reflects a mode of thinking in which a connection, a perception of sameness along some dimension (semantic, formal, phonic, etc.), is made between two linguistic units; changes caused by such influence are referred to as analogical changes and while a number of classificatory schemata are possible for the variety of attested analogical changes,[26] virtually all of these changes boil down to the same

basic motivation, that of echoing the abovementioned perception of sameness by the construction of a sameness in form. For instance, in the change of the Greek second person singular past ending, from *-so* to *-sun*, it appears that there was influence of (i.e., a perception of sameness with) the first person singular ending *-mun*, since in this case, there was no general change of *o* to *u* nor a general accretion of a word-final *-n* that could have altered the earlier *-so* to *-sun*. Moreover, the grammatical closeness of the endings in terms of what they mark on a verb makes an appeal to analogical influence particularly attractive here. Further, the change mentioned above of early Latin *paros* 'small' to later *parwos*, the mismatch between a stem form *par-* in the early nominative singular and a stem form *parw-* in the genitive singular suggests that the change to both forms having *parw-* shows a similar motivation; a clear connection between the two – they are members of the same paradigm after all – can be taken as the basis for the influence of one form (here the genitive form) over another (here the nominative form) and the formal reshaping of the latter in accordance with this influence. The psychological/cognitive link between the forms, here furnished by their grammatical sameness, provides the basis for the change. Even in the case of the generalization of meaning and semantic reinterpretations of the sort seen with *dogge* → *dog*, psychological factors play a role, since in a sense the changes represent reasonable guesses as to the connection between a word and the context it occurs in; that is, since even two animals of the same species are not point-for-point identical in all respects (trivially, they can differ in size and age), a speaker hearing *dogge* being used to refer to two separate canines, even ones ostensibly similar in some respects, could make the reasonable assumption that the word could be used in the case of *any* canine nonidentity – that is, such an assumption would be an instance of an abductive change, in the sense of Andersen (1973), motivated by a reasoning schema involving a "best guess" as to what the use of a particular word was focusing on. Finally, to the extent that universals of linguistic structure and use can be identified that have some reasonable cognitive basis, some changes can be attributed to such cognitive factors; the change in (2) above in which Greek came to require an object pronoun in a construction that previously did not require it may be a case in point, if a perceptually based universal constraint that favors finite clauses that are whole and intact, as opposed to the "streamlining" possible with reduced clauses such as infinitives, is responsible for the appearance of the object pronoun in the later Greek form of the construction (as suggested tentatively in Joseph 1980, though see Joseph 1990: 186–7, 197n.B, 201–2 for some counterindications).

One way of telling that a psychological cause such as analogy is responsible for a change is that other causal factors can be ruled out. In particular, there is no reason to think that physiological factors, such as the constraints of the speech tract or the perceptual mechanism, a type of explanation pursued very compellingly by Ohala (see, e.g., Ohala 1993, 2003), were at work. Still, in most cases of pure sound change physiology does play a leading role. The very common loss of unaccented, especially unstressed, vowels (see (3d)), can be attributed to the weak articulation of an unaccented vowel when the main accent involves heightened intensity (as it does in English), though the weak perceptual salience of such vowels plays a role too. Moreover, assimilation (see (3e)), surely the single most common type of sound change there is, is triggered mostly by the greater economy of articulator movements needed in the transition from one sound into the next when the sounds agree, e.g., in point of articulation (as in (3e)).

In a sense, both analogy and physiologically induced sound changes involve aspects of the language system as a system. Analogy, for instance, pertains in part to the mental storage of linguistic material or the cognitive side thereof, and has to do as well with the systems of relations among elements that speakers perceive and establish. Physiology, moreover, pertains to those parts of the system involved in the production or perception of speech. Still, there are other system-related factors that play a role in bringing on language change. Some of the shifts in long vowels seen in English, for instance, were not isolated events but rather were tied to other changes in the vowel system; thus, (roughly) not only did mid front ē become ī (as in *wē* to Modern *we*, discussed earlier) but also low ā became ē (as in *name*). Such "chain shifts" seem to involve

whole systems of sounds moving rather than there being a series of completely isolated and unrelated changes. Similarly, the crowding of phonetic space referred to above (Section 2) concerning Serbian affricates would be a clear case of systemic pressures playing a role in a change in those dialects that have narrowed the original three-way contrast to a two-way one (see note 7). Finally, at the lexical level, one can note the so-called "blocking effect" where the existence of a fixed expression in a language seems to be able to block the creation of synonymous expressions, so that the system of lexical(ized) expressions interacts with the productive mechanisms for spontaneous creation of lexical material; thus the presence of *yesterday* in English apparently blocks the phrase **the day before today*, whereas the absence of a word like **pre-yesterday* conversely seems to play a role in the acceptability of the phrase *the day before yesterday*.

Finally, there are social factors that play a role in causing language change. Some matters in language are directly sensitive to speakers' place in society and their relationship with other speakers, in particular terms of address and epithets; when there are changes in the social fabric, there can be corresponding changes in these linguistic aspects, usually involving lexical change. For instance, during the period around the French Revolution, changes took place in the form of second person address in French, in accordance with a general egalitarian ideology in which the reciprocal use of the ("familiar") second singular pronoun *tu* served as an expression of solidarity.[27] Similarly, changes in attitudes about various sorts of designated groups in American society have led to changes in their appellations, giving, for instance, *differently abled* instead of *handicapped*, *First (*or *Native) Americans* instead of *Indians*, etc.

There are, however, other, perhaps more important, ways in which social factors play a role in change, for they provide the key mechanism for the spread of one of a set of competing forms throughout a speech community, largely through the attachment of prestige to one variant. As noted in Section 2, both the "embedding" problem and the "evaluation" problem involve the recognition of language as a quintessentially social phenomenon, and the evaluation problem is especially relevant to the matter of the spread of innovations. The use of language as a marker of social identity and group membership means that various aspects of language use can spread among members of a group, if – for whatever reason – these features are taken to be emblematic of individuals identified as key or typical members of a group. This process can be seen, for instance, in the spread of slang expressions or jargon (i.e., occupationally related vocabulary), where one's "in-group" versus "out-group" status based on use of or knowledge of particular terms and phrases is often painfully evident, as any older speaker in the midst of a group of teenagers or a nonenthusiast amongst a group of "technophiles" can readily attest to. Importantly, the same mechanisms that foster the spread of such lexical innovations seem to be at work in more subtle kinds of change involving innovative pronunciations, constructions, and the like. Admittedly, though, it is still an unresolved issue among linguists as to when one can talk about a change – at the point at which an innovation arises, e.g., due to systemic or physiological factors, as outlined above, or at the point at which an innovation has spread, having been adopted by speakers beyond the point of origination.[28]

The recognition of the role of social factors leads to one particular type of social situation involving speakers of a language, namely when they come into contact with speakers of a different language. Such language contact situations are in a sense no different in kind from the contact between speakers of different dialects of the same language, though the degree of difference between the speech forms exhibited by each speaker is typically greater in the case of language contact. Language contact can be the source of innovations, most evidently in lexical matters. For example, new words or phrases can enter a language from models in another language, in the form of direct borrowings such as *praise*, borrowed into Middle English from early French and ultimately replacing earlier English *herian* (cf. (1a) above), and *coup d'état*, more recently borrowed, also from French, but also via so-called "loan translations" in which a foreign phrase is rendered into the borrowing language, as with the phrase *It goes without saying*, based almost literally on French *Ça va sans dire*. Sometimes, however, borrowings can directly or indirectly introduce

structural innovations into a language. For example, the existence of plurals in English such as *schemata* or *criteria* or *bases* (from *basis*), all from Greek, has extended the range of plural formation possibilities, and has led to innovative forms such as *processes*,[29] pronounced with a final syllable [… siyz]), modeled analogically on *bases*; similarly, the active voice *-ing* form in *it goes without saying* is unusual from the English standpoint, where a passive form as in *it goes without being said* would be, strictly speaking, more "English-like."

Under intense conditions of sustained language contact, especially when there is some degree of bi- or multilingualism to be found among individuals in a speech community,[30] it is not unusual for languages to converge structurally. This has happened in the Balkans, where Albanian, Bulgarian, Greek, Macedonian, and Romanian, among other languages, have come to be syntactically quite parallel to each other, so much so that the languages have been spoken about as multiple lexicons with but a single grammar.[31] The social context in which contact takes place turns out to have a significant effect on the outcome of the linguistic contact, to the extent that the current thinking is that there are no linguistic constraints whatsoever on what may be transferred from one language into another in a contact situation – one finds all types of words and morphemes borrowed, sentence patterns passing between languages, meanings of words being affected, new sounds entering a language, and so on, all through contact.[32]

The effects of contact are so pervasive, especially when one considers that the spread of innovations within a language necessarily involves contact among speakers, in such a case though of the same language, as noted above, that it could be hypothesized that *all* change in language involves contact. Despite the potential for such a claim, the noncontact causes of change, outlined above, cannot be discounted, and it seems that the causes of language change are best understood by reference to both internal and external factors.

6 Some Dramatic Discoveries and Important Methods

This survey of historical linguistics would be incomplete without mention of two dramatic discoveries among the many that have emerged from this subfield: language relatedness and regularity of sound change. These discoveries also have the benefit of allowing for a consideration of certain key methods that historical linguists have utilized over the years.

With regard to the former, we observe that scholars have long been intrigued by the mix of diversity and similarity that human languages show. Among the hypotheses that have been advanced to explain this mix, among the most promising is one that claims that at least some of the known languages show certain similarities because they represent later instantiations of a once single speech community; that is, it has been hypothesized that a single speech community, through the accumulation of changes of the sort described in previous sections and perhaps aided by migrations, resettlement, and physical splits in the community, can over time divide and spawn numerous separate and ultimately distinct speech communities. In such a situation, the resulting distinct speech communities show some similarity by virtue of deriving from the same starting point, and more important, show various systematic correspondences of form for this same reason. These resulting languages are said to be *related* (actually, *genetically related*, where "genetic" has its etymological sense of "pertaining to origin" not the more modern, biological, sense),[33] and the original speech community is referred to as a *protolanguage* (or *parent language*) for its several offspring languages.

The recognition that languages could be viewed as related to one another, led, by extension, to the observation that some languages were more closely related to each other than to other languages. Such clusters of more closely related languages are said to form *subgroups* within a larger language family. With that recognition, therefore, grouping and subgrouping of languages became an important scholarly activity, and with the discovery of new languages, the question of how they fit into the emerging set of known language families needs always to be asked.

Critical to the establishment of relatedness is the issue of methodology.[34] Of paramount importance here is the comparative method, by which corresponding features (more usually sounds in corresponding words but also morphemes and even syntactic structures) are compared with an eye to determining a set of systematic relationships that hold among the languages being compared. Languages are generally held to be related when a sufficiently large set of such correspondences can be found, though there are controversies over just how large such a set needs to be to warrant a claim of relatedness, and whether the correspondences could instead be a matter of chance or perhaps due to contact between the languages in question. When such systematic correspondences can be found, then one can also draw inferences about the source from which the related languages arose, on the assumption that the comparable elements each derived through their own lineal descent from a common starting point. When the comparative method "works," therefore, it is possible to make hypotheses about the earlier states from which the related languages developed and thus to reconstruct (aspects of) ancestor languages that gave rise to the set of related languages in question. For example, the recurring correspondence set described below involving *p* in Greek, Latin, and Sanskrit matching *f* in Germanic (under certain conditions), has led most Indo-Europeanists to a reconstruction of *p* for the sound in the source language ("Proto-Indo-European") that gave rise to the corresponding elements in the offspring languages.

A side benefit for the study of language change is the fact that the assumption of relatedness and the comparative method also provide another source of information about change. If an element A in one language can be systematically compared to a nonidentical element B in another (putatively related) language, and the hypothesis is made that they derive from a reconstructed element C (usually affixed with a * to indicate that the reconstruction is a hypothesis not an attested form), then clearly at least one change has occurred – either A has changed and B reflects the reconstructed element faithfully, or B has changed and A has not, or both A and B have changed, in different directions). Thus if we reconstruct Proto-Indo-European *p for the set of Sanskrit (etc.) *p* = Germanic *f*, we are committing ourselves to the hypothesis that Germanic is innovative in this case; had we reconstructed something like an affricate *pf, then we would be committed to that view that there had been change in all the languages being compared (*pf > p in some and *pf > f in others).

As a result of all the research into language relatedness and grouping of languages into families, there are now numerous well-researched and well-established language groups. Among these, to name just a few, are *Indo-European* ("IE," covering many of the languages from India west into Europe, including English, French, Greek, Russian, among numerous others), *Finno-Ugric* (covering Hungarian and many languages in the Baltic area, including Estonian and Finnish), *Sino-Tibetan* (including Tibetan, Burmese, and the numerous Chinese languages, Mandarin, Cantonese, etc.), *Semitic* (taking in languages of the Middle East, including Hebrew, Arabic, and ancient Akkadian), *Bantu* (covering numerous languages of eastern and southern Africa, such as Swahili, Setswana, and Zulu), *Algonquian* (including many native North American languages from the eastern seaboard across the Great Lakes area into the prairie provinces of Canada, such as Cree, Fox, Ojibwa, Micmac, Massachusett, Delaware, etc.), *Uto-Aztecan* (covering a huge number of languages of the western United States and Mexico, including Comanche, Southern Paiute, Hopi, Nahuatl, and others), *Athabaskan* (covering languages extending from Alaska into Mexico, including Chipewyan, Navajo, and Apache), and *Austronesian* (covering much of the South Pacific, including Tahitian, Samoan, Maori, Hawaiian, and Tagalog, but extending also into Madagascar where Malagasy is spoken). There are also several languages that have defied classification and so are called *language isolates*, e.g., Basque, spoken now in southern France and northern Spain; Burushaski, still spoken in the northern part of South Asia; and Sumerian, spoken in ancient times in Mesopotamia. Such languages have no known or demonstrable relatives, though it is conceivable, even likely, that they have relatives that are no longer spoken, i.e., that died out without a trace, or relatives that current methods simply are not able to link to the isolates with any degree of certainty (and see below).

Some of these groups are widely recognized to be themselves part of still larger, more all-compassing groupings. For instance, Finno-Ugric is considered to be part of the Uralic family (covering various languages in Siberia, e.g., the Samoyed languages east of the Ural mountains); Semitic is held to be part of Afro-Asiatic (covering (Ancient) Egyptian, Berber, Hausa, and others); Bantu is seen to be part of Niger-Congo (covering West African languages such as Yoruba, Igbo, Twi, and others); Algonquian is taken to be related to two now extinct languages in California (Wiyot and Yurok) and thus to be part of a larger, so-called Algonquian-Ritwan or Algic, family; and so on.

These well-recognized larger groupings raise interesting questions, and ongoing controversies, regarding the extent to which all languages can be shown to fall into ever larger groupings. Is Indo-European related to Uralic, as many believe, and to Semitic? Do these families cohere as part of an even larger so-called Nostratic family, covering as well other families such as Kartvelian (in the Caucasus), Altaic (in Central and Eastern Asia), etc.? Does Austronesian form a larger grouping with Sino-Tibetan? Do the numerous language families in North and South America show any further groupings, perhaps into as few as two or three megafamilies? More generally, how far can such "lumping" of languages go? In particular, can a single protolanguage be posited for all known languages?[35]

Armed with these hypotheses about relatedness, linguists in the nineteenth century, especially Western European scholars investigating the Indo-European languages, were struck by the discovery of numerous systematic correspondences of sounds in various languages in Europe and Asia believed to be part of the IE family, and eventually also by their ability to formulate these correspondences in a precise way, so that apparent exceptions to the correspondences turned out to be systematic in their own right. For instance, the Danish scholar Rasmus Rask (1818) and the German polymath Jacob Grimm (1819) described various correspondences that held between stop consonants in Sanskrit, Greek, Latin, and Germanic, e.g., as noted above, Skt. *p* = Grk. *p* = Lat. *p* = Gmc. *f*, but also *d* = *d* = *d* = *t*, with both correspondences seen in *pad-* = *pod-* (ποδ-) = *ped-* = *foot*. Moreover, many instances of these sets, and others like them involving other points of articulation, were brought to light. Exceptions to these sets were found too, though, yet they were soon explained; for instance, Skt. *p* = Gmc. *p* in Skt. *spaś-* = Old High German *spehon* 'see' or Lat. *p* = Gmc. *p* in *spuo* = *spit*, were shown by Carl Lottner (1862) to occur only after *s*, and cases such as Skt. *p* = Grk. *p* = Lat. *p* = Gmc. *v*, as in *saptá* = *heptá* (ἑπτά) = *septem* = *seven*, where Germanic showed a voiced fricative, were shown by Karl Verner (1877) to be conditioned by the original position of the word accent, since the p/p/p/f set occurs before the accent while p/p/p/v set occurs after the accent, taking the Sanskrit and Greek accent to be indicative of its original placement in Proto-Indo-European (and thus in pre-Germanic).

Successes such as these, and others, meant that all of the exceptions to Grimm's observations could be accounted for in a systematic way. The result was that the sound correspondences could be said to be *regular*, in that they held for sounds meeting particular linguistic conditions, e.g., the nature of adjacent sounds, the position relative to accent, etc., conditions which really defined subregularities in their own right. The empirical claim that emerged from such observations was that sound change was regular, subject only to precisely formulable phonetic conditioning. The exceptionlessness of sound change became an important rallying point for historical linguists in the nineteenth century, and this hypothesis, often referred to now as the Neogrammarian view of sound change, after the scholars based mostly in Leipzig who advanced this notion most vigorously, put the field of linguistics on a scientific footing. Holding only phonetic factors responsible for sound change meant that sound change could be seen as triggered essentially only by physiological factors, of the sort discussed in Section 5. The Neogrammarian assumptions about sound change have generally withstood the test of time and the challenges of careful examination of case after case of sound change from language after language and continue to have importance in linguistics today; for instance, it is not unreasonable to see the insistence in generative grammar (see Chapters 7 and 13) on rule-governed aspects of language as an outgrowth of the Neogrammarian discovery of the regularity of sound change.

7 For the Future: What Remains to Be Done?

It should be clear that much has been accomplished toward understanding what happens to languages through time, the basic subject matter of historical linguistics. But even with these impressive accomplishments, much still remains to be done.

First, for all that is known about the histories of numerous individual languages, there are still many languages whose history has not been investigated carefully. In some instances, such investigation is a matter of mining the available material, e.g., regarding Medieval Greek, or Albanian after the sixteenth century, while for others it involves working out or exploring further relatedness relations with other languages and using the comparative method and/or other methods to make inferences about the earlier stages of the language in question.

Even for well-researched languages, more cataloging of changes, as well as the determination of a myriad of details of developments, is needed; many texts remain underexamined from all stages of even a language such as English and the same holds for Greek, French, Russian, Hindi, and so on and so forth. Here, what is needed also is information about the social setting for all these languages at all relevant stages, in keeping with the "embedding" problem referred to in Section 2.

Besides filling the gaps in language history, such further research will help toward the development of a clear characterization of naturalness, and thus feed into the development of a general theory of language change, another desideratum that at present eludes us, as the discussion in Section 2 of the "constraints" problem indicates.

With regard to relatedness among languages, it is fair to ponder whether we have hit a ceiling beyond which there is no further progress. The questions posed at end of the discussion in the previous section are thus directions for future research but are perhaps ultimately unanswerable. It is worth observing here that, as inherently interesting as these questions are, even if they could be answered, even if a "protoworld" could be confidently posited, there would still be the question of how the diversity evident in the languages of the world arose. That is, remaining issues of relatedness are only part of what remains to be done in historical linguistics. Some promising and very intriguing results have been obtained in recent years by the use of computational phylogenetic methods taken from evolutionary biology, but there is considerable controversy.[36]

Moreover, what may be thought of as the ultimate historical linguistic question of the origin of language still awaits a definitive answer, and may never be resolved. See note 35 but especially Chapter 1 for some discussion.

Finally, putting together all the research on language change and historical linguistics leads one to wonder whether a general theory of change is possible. Here it must be recognized that such a theory would involve working out the parameters of change, essentially answering the five key questions in Section 2, but paying attention as well as to diachronic stability, for it is not the case that everything in a language necessarily will undergo a change.[37]

8 Conclusion

Of necessity, this survey has not been able to provide detail on all matters that make up the subfield of historical linguistics, but one final important point is that, in order to do historical linguistics properly, one needs above all else to be able to handle all sorts of subfields of linguistics properly. A full understanding of the synchronic system of a language at (at least) two different stages is essential to understanding what has changed and what has not; sociolinguistics must be invoked in order to fully understand the context in which changes occur and especially spread; phonetics is relevant to understanding sound change; and so on. Thus while not in the center of the field of linguistics,[38] historical linguistics nonetheless draws on virtually all aspects of the field in ways that other subfields do not.

NOTES

1 See Ohala (1980, 1994: 332–5) on the possible origins of smiling and thus its functional stability over the ages.

2 See Schourup (1985 [1982]) for an early discussion of this innovative use of *be like*, which now has a huge literature, both synchronic and diachronic, associated with it; among studies treating it from the perspective of diachrony are Romaine and Lange (1991); Blyth, Recktenwald, and Wang (1995); and Cukor-Avila (2002). Butters (1980) discusses the extent to which the narrative use of *go* was itself an innovation earlier in the twentieth century.

3 The translation is taken from Fowler (1977).

4 The translation here is based on the text and notes in Shoaf (1989). I thank Bethany Christiansen for help with details of this translation.

5 See Janda (2001) and Janda and Joseph (2003) for discussion.

6 See Melchert (1991) for a particularly moving account of the universality of a Hittite king's fears when facing death; Joseph (1998) gives a classroom application of Melchert's insights.

7 As it happens, many Serbian speakers innovatively do not have the older three-way distinction anymore, so some mergers have occurred here. My thanks to Ronelle Alexander of the University of California, Berkeley for clarification of this point.

8 Fodor (1965) has some very interesting, but ultimately inconclusive, discussion on this issue.

9 Note that ē → e and ē → a are changes that are attested in other languages (e.g., the former in Pontic Greek, the latter in Bulgarian (with a palatal on-glide), and are thus possible outcomes of change that one has to reckon with (though it is not clear if these are direct changes or the result of the accumulation of several changes). For a discussion of why vowels move along the paths they do, see Labov (1994), especially the Appendix.

10 For instance, the use of the marker *tou* (originally a genitive case form of the definite article used as a nominalizer of verbs) as a generalized complementizer introducing the subordinated infinitive disappears from later Greek (compare the reduction in English from the double complementizer of Chaucerian *whan that* to the later single complementizer discussed above). Similarly, the status of the marker *na* has changed; it was most likely a full-fledged complementizer when it was first used as a generalized subordinator in Medieval Greek (it derives from the Ancient Greek final conjunction *hína* 'so that') but in Modern Greek it is arguably merely a grammatical marker of the subjunctive mood (see Philippaki-Warburton 1994).

11 This is the phenomenon known as Watkins' Law (Watkins 1962), discussed with additional references in Collinge (1985: 239–40); see also Janse (2009).

12 This is the observation embodied in Kurylowicz's fourth "law" of analogy (Kurylowicz 1947); see Hock (1991: 210–37) for discussion, and Winters 1995 for a translation and added discussion.

13 This phenomenon is referred to in the literature as "grammaticalization" (sometimes also "grammaticization" or even "grammatization"); see Hopper and Traugott (2003) and Heine (2003) for an introduction to the study of such phenomena, which has engendered an enormous literature; Narrog and Heine (2011) explores grammaticalization from numerous angles, while Heine and Kuteva (2002) offers a wide range of examples of grammaticalization from numerous languages. See Campbell (2001); Janda (2001); Joseph (2001, 2011a, 2014); Newmeyer (1998), and Norde (2001) for some critical reappraisals of some of the claims of so-called grammaticalization theory.

14 Such a development is known in the literature as "degrammaticalization," among other similar designations (e.g., "antigrammaticalization"). See Janda (2001) for a summary of the rather considerable number of such cases that have been documented. An in-depth discussion of degrammaticalization, along with other examples, can be found in Norde (2009).

15 There are recurrent patterns in semantic change, however; see Traugott and Dasher (2002) for a presentation of semantic shifts that are commonly occurring (giving what they call "regularity").

16 I say this to remind the reader that language is not exclusively a matter of the vocal channel, since manually based sign(ed) languages are full-fledged languages in all respects known to us. From a diachronic perspective, sign(ed) languages show many of the same types of change as vocally based languages do, and their users respond to the same types of social factors that affect change in all languages. See Frishberg (1975, 1976); Fischer (2014); and Hock and Joseph (2009: 129, 130–1, 166, 256–7) for some examples and discussion.

17 See King (1969) for a summary of these views in a (then) definitive statement, and Jasanoff (1971) for a highly critical assessment of them.

18 See, for instance, Nagy and Reynolds (1997).

19 Note that the view that grammar change is motivated by simplicity alone could use the number of rules as a metric for evaluating the simplicity of a grammar.
20 Compare also current versions of Optimality Theory where the constraints that are ranked are output-oriented.
21 Thus *male* "contaminated" *femelle* and a blended form *female* resulted.
22 See Joseph (1994, 2000, in prep.), for more details on this development.
23 See Lloyd (1987: 162ff.), and Penny (1991: 150ff.) for some discussion. I am indebted to Rich Janda for bringing this example to my attention.
24 See Labov (1994) for an excellent and detailed survey of the results of this research program into the spread of change.
25 See Joseph (2012) on the nature of diffusion in general, and Labov (2001, 2010) for more on the processes of the spread of innovations.
26 See the discussion and presentation of terminology in virtually any standard textbook on historical linguistics, e.g., Hock (1991) or Hock and Joseph (2009), among (many) others.
27 See Brown and Gilman (1960) for a discussion of these and other developments pertaining to second person address in various European languages.
28 See Joseph (2011b) for some discussion of this very question.
29 The noun *process* is a borrowing ultimately from Latin, and thus a Greek-like plural would not be expected with it; once it enters English, of course, all bets are off, and the word is no longer bound by its heritage. Attaching the native English plural marker or a Greek-like marker or reanalyzing the word are all within the realm of possibility; note that *criteria* is quite commonly used as a singular by many American English speakers, and one can even occasionally hear *criterions*.
30 Recognizing the role of multilingualism in language change brings a seemingly "external" cause, namely language contact, into the "internal" – here psychological – domain, since the "contact" is really in the mind of the bilingual speaker.
31 This quote is based on the observation of the Slovene linguist Kopitar who noted (1829: 86) concerning Albanian, Bulgarian, and Romanian that "nur eine Sprachform herrscht, aber mit dreierlei Sprachmaterie" ('only one grammar holds sway, but with three lexicons'). I follow here the translation by Friedman (1999: 3), who has very insightful comments to make about the Balkan speech community.
32 See Thomason and Kaufman (1988: ch. 2) and Thomason (2003) for discussion of this point.
33 Some linguists now use "genealogical" instead of "genetic" here so as to avoid interference from the biological sense of "genetic."
34 See Campbell and Poser (2008) for discussion of the methodology of determining language relatedness.
35 To some extent, therefore, such questions can lead into speculation about the ultimate origin of language (see Chapter 1) – if human language originated in a single place, then a "protoworld" might be conceivable, though most likely not reconstructable, but if language arose independently in various places around the world, then a "protoworld" could not be a coherent notion. Overlooked in much of the debate and speculation about a "protoworld" (though see the brief comments in Hock and Joseph 2009: 458, 466) and Salmons and Joseph 1998: 3n.7) is the fact that numerous fully natural and complex sign(ed) languages have arisen spontaneously around the world in various communities with significant numbers of deaf people, so that at best, it would seem that "protoworld" is "proto-oral-world" and not a protolanguage for *all* natural human languages.
36 See Dunn (2014) and François (2014) for excellent surveys of some of these methods and the results they have gotten. A particularly interesting and occasionally somewhat caustic exchange focusing on these methods as applied to the question of the origin and spread of Austronesian can be seen through Gray, Drummond, and Greenhill (2009); Donahue, Denham, and Oppenheimer (2012a); Greenhill and Grey (2013); and Donahue, Denham, and Oppenheimer (2012b).
37 See Nichols (2003) for a discussion of some aspects of language that show stability through time.
38 One might qualify this statement with the modifier "anymore," for, in the nineteenth century, diachronic linguistics *was* linguistics, period.

REFERENCES

The bibliography on historical linguistics is vast, and the works specifically referred to here do not even begin to cover the field. For reasonably good bibliographic coverage, relatively recent

textbooks such as Campbell (2013), Hock (1991), Hock and Joseph (2009), Trask (1996), Millar (2013), Ringe and Eska (2013), among others, should be consulted; see also handbooks such as Joseph and Janda (2003), Bowern and Evans (2014), Janda, Joseph, and Vance (2017), for up-to-date surveys of the field at large.

Andersen, H. (1973). Abductive and deductive change. *Language*, 49, 765–93.

Blyth, C., Jr., Recktenwald, S., and Wang, J. (1990). I'm like, "say what?!": A new quotative in American oral narrative. *American Speech*, 65, 215–27.

Bowern, C. and Evans, B. (2015). *The Routledge Handbook of Historical Linguistics.* London: Routledge.

Brown, R. and Gilman, A. (1960). The pronouns of power and solidarity. In T. Sebeok (ed.), *Style in Language* (pp. 253–76). New York: John Wiley & Sons, Inc.

Butters, R. (1980). Narrative go "say". *American Speech*, 55, 304–7.

Campbell, L. (2001). What's wrong with grammaticalization? *Language Sciences* (Special Issue, *Grammaticalization: A Critical Assessment*, edited by L. Campbell), 23, 2–3.

Campbell, L. (2013). *Historical Linguistics: An Introduction*, 3rd edn. Edinburgh: Edinburgh University Press.

Campbell, L. and Poser, W. (2008). *Language Classification: History and Method*. Cambridge: Cambridge University Press.

Collinge, N. E. (1985). *The Laws of Indo-European.* Amsterdam: John Benjamins.

Cukor-Avila, P. (2002). *She say, she go, she be like*: Verbs of quotation over time in African American vernacular English. *American Speech*, 77, 1, 3–31.

Donohue, M., Denham, T., and Oppenheimer, S. (2012a). New methodologies for historical linguistics? Calibrating a lexicon-based methodology for diffusion vs. subgrouping. *Diachronica: International Journal for Historical Linguistics*, 29, 4, 505–22.

Donohue, M., Denham, T., and Oppenheimer, S. (2012b). Consensus and the lexicon in historical linguistics: Rejoinder to "Basic vocabulary and Bayesian phylolinguistics". *Diachronica: International Journal for Historical Linguistics*, 29, 4, 538–46.

Dunn, M. (2014). Language phylogenies. In Bowern and Evans (2014), pp. 190–211.

Fischer, S. D. (2014). Sign languages in their historical context. In Bowern and Evans (2014), pp.442–65.

Fodor, I. (1965). *The Rate of Linguistic Change. Limits of the Application of Mathematical Methods in Linguistics.* Janua Linguarum, Series Minor, 43. The Hague: Mouton.

Fowler, H. N. (trans.) (1977). *Plato in Twelve Volumes. IV: Cratylus, Parmenides, Greater Hippias, Lesser Hippias.* Loeb Classical Library, 167. Cambridge, MA: Harvard University Press.

François, A. (2014). Trees, waves and linkages: models of language diversification. In Bowern and Evans (2014), pp. 161–89.

Friedman, V. (1999). *Linguistic Emblems and Emblematic Languages: On Language as Flag in the Balkans.* Kenneth E. Naylor Memorial Lecture Series in South Slavic Linguistics (edited by Brian D. Joseph), 1. Columbus, OH: Department of Slavic and East European Languages and Literatures, The Ohio State University.

Frishberg, N. J. (1975). Arbitrariness and iconicity: historical change in American Sign Language. *Language*, 51, 696–719.

Frishberg, N. J. (1976). *Some Aspects of the Historical Development of Signs in American Sign Language.* PhD dissertation, University of California, San Diego.

Gray, R. D., Drummond, A. J., and Greenhill, S. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*, 323, 479–83.

Greenhill, S. J. and Gray, R. D. (2012). Basic vocabulary and Bayesian phylolinguistics. *Diachronica: International Journal for Historical Linguistics*, 29, 4, 523–37.

Grimm, J. (1819). *Deutsche Grammatik.* Göttingen: Dieterich.

Halle, M. (1962). Phonology in Generative Grammar. *Word*, 18, 54–72.

Heine, B. (2003). Grammaticalization. In Joseph and Janda (2003), pp. 575–601.

Heine, B. and Kuteva, T. (2002). *World Lexicon of Grammaticalization.* Cambridge: Cambridge University Press.

Hock, H. H. (1991). *Principles of Historical Linguistics*, 2nd edn. Berlin: Mouton de Gruyter.

Hock, H. H. and Joseph, B. D. (2009). *Language History, Language Change, and Language Relationship. An Introduction to Historical and Comparative Linguistics*, 2nd edn. Berlin: Mouton de Gruyter. (1st edn 1996.)

Hopper, P. J. and Traugott, E. C. (2003). *Grammaticalization*, 2nd edn. Cambridge: Cambridge University Press. (1st edn 1993.)

Janda, R. D. (2001). Beyond 'pathways' and 'unidirectionality': on the discontinuity of language transmission and the counterability of grammaticalization. *Language Sciences* (Special Issue, *Grammaticalization: A Critical Assessment*, edited by L. Campbell), 23, 2–3.

Janda, R. D. and Joseph, B. D. (2003). On language, change, and language change. In Joseph and Janda (2003), pp. 3–180.

Janda, R. D., Joseph, B. D., and Vance, B. (eds) (2017). *Handbook of Historical Linguistics*, Vol. 2. Malden, MA: Wiley Blackwell.

Janse, M. (2009). Watkins' Law and the development of agglutinative inflections in Asia Minor Greek. *Journal of Greek Linguistics*, 9, 93–109.

Jasanoff, J. (1971). Review of King 1969. *Romance Philology*, 25, 74–85.

Joseph, B. D. (1980). Linguistic universals and syntactic change. *Language*, 56, 345–70.

Joseph, B. D. (1990). *Morphology and Universals in Syntactic Change: Evidence from Medieval and Modern Greek*. New York: Garland Publishing. (Updated and augmented version of 1978 Harvard University PhD Dissertation.)

Joseph, B. D. (1994). On weak subjects and pro-drop in Greek. In I. Philippaki-Warburton, K. Nicolaidis, and M. Sifianou (eds.), *Themes in Greek Linguistics (Papers from the First International Conference on Greek Linguistics, Reading, September 1993)* (pp. 21–32). Amsterdam: John Benjamins.

Joseph, B. D. (1998). Linguistics for Everystudent. *Studies in the Linguistic Sciences*, 28, 2, 123–33.

Joseph, B. D. (2000). Processes of Spread for Syntactic Constructions in the Balkans. In C. Tzitzilis and C. Symeonidis (eds.), *Balkan Linguistik: Synchronie und Diachronie* (pp. 139–50). Thessaloniki: Aristotle University of Thessaloniki.

Joseph, B. D. (2001). Is there such a thing as "grammaticalization"? *Language Sciences* (Special Issue, *Grammaticalization: A Critical Assessment*, edited by L. Campbell), 23, 2–3, 163–86.

Joseph, B. D. (2011a). Grammaticalization: a general critique. In Narrog and Heine (2011), pp. 193–205.

Joseph, B. D. (2011b). Historical linguistics and sociolinguistics – strange bedfellows or natural friends? In N. Langer, S. Davies and W. Vandenbussche (eds), *Language and History, Linguistics and Historiography* (pp. 67–88). Oxford: Peter Lang.

Joseph, B. D. (2012). Lexical diffusion and the regular transmission of language change in its socio-historical context. In J. M. Hernández-Campoy (ed.), *The Handbook of Historical Sociolinguistics* (pp. 408–26). Oxford: Wiley Blackwell.

Joseph, B. D. (2014). What counts as (an instance of) grammaticalization? *Folia Linguistica* (Special Issue, *Refining Grammaticalization*, edited by F. von Mengden and S. Horst), 48, 361–83.

Joseph, B. D. (in prep.). The Modern Greek Weak Subject Pronoun τος – Its Origins and Implications for Language Change and Language Structure: A Study in Grammatical Change. To appear in Studies in Diachronic Syntax Series, Oxford: Oxford University Press.

Joseph, B. D. and Janda, R. D. (eds). (2003). *Handbook of Historical Linguistics*. Oxford: Blackwell.

King, R. (1969). *Historical Linguistics and Generative Grammar*. Englewood Cliffs, NJ: Prentice Hall.

Kiparsky, P. (1968). Linguistic universals and linguistic change. In E. Bach and R. T. Harms (eds), *Universals in Linguistic Theory* (pp. 171–202). New York: Holt, Rinehart & Winston.

Kiparsky, P. (1971). Historical linguistics. In W. O. Dingwall (ed.), *A Survey of Linguistic Science* (pp. 576–649). College Park, MD: University of Maryland.

Klima, E. (1964). Relatedness between grammatical systems. *Language*, 40, 1–20.

Klima, E. (1965). *Studies in Diachronic Transformational Syntax*. PhD dissertation, Harvard University.

Kopitar, J. (1829). Albanische, walachische und bulgarische Sprache. *Jahrbücher der Literatur*, 46, 59–106.

Kurylowicz, J. (1947). La nature des procès dites analogiques. *Acta Linguistica*, 5, 15–37.

Labov, W. (1982). Building on empirical foundations. In W. Lehmann and Y. Malkiel (eds), *Perspectives on Historical Linguistics* (pp. 17–82) (= Current Issues in Linguistic Theory, 24). Amsterdam: John Benjamins.

Labov, W. (1989). The child as linguistic historian. *Language Variation and Change*, 1, 85–94.

Labov, W. (1994). *Principles of Linguistic Change. Vol. 1: Internal Factors*. Oxford: Blackwell.

Labov, W. (2001). *Principles of Linguistic Change. Vol. 2: Social Factors*. Malden, MA: Blackwell.

Labov, W. (2010). *Principles of Linguistic Change*. Vol. 3: *Cognitive and Cultural Factors*. Malden, MA: Wiley-Blackwell.

Lloyd, P. M. (1987). *From Latin to Spanish. Vol. I: Historical Phonology and Morphology of the Spanish Language*. Memoirs of the American Philosophical Society, 173. Philadelphia, PA: American Philosophical Society.

Lottner, C. (1862). Ausnahmen der ersten Lautverschiebung. *Zeitschrift für vergleichende Sprachforschung*, 11, 161–205. (Translated in W. Lehmann (ed.), *A Reader in 19th Century Historical Indo-European Linguistics*. Bloomington, IN: Indiana University Press.)

Melchert, H. C. (1991). Death and the Hittite king. In *Perspectives on Indo-European Language, Culture and Religion. Studies in Honor of Edgar C. Polomé*, Vol. I (pp. 182–8). *Journal of Indo-European Studies*, Monograph No. 7. McLean, VA: Institute for the Study of Man.

Millar, R. M. (2013). *Trask's Historical Linguistics*. London: Routledge. (2nd (revised) edition of Trask 1996.)

Nagy, N. and Reynolds, B. (1997). Optimality theory and variable word-final deletion in Faetar. *Language Variation and Change*, 9, 1, 37–55.

Narrog, H. and Heine, B. (eds.) (2011). *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press.

Newmeyer, F. J. (1998). *Language Form and Language Function.* Cambridge, MA: MIT Press.

Nichols, J. (2003). Diversity and stability in language. In Joseph and Janda (2003), pp. 283–310.

Norde, M. (2001). Deflexion as a counterdirectional factor in grammatical change. *Language Sciences* (Special Issue, *Grammaticalization: A Critical Assessment*, edited by L. Campbell), 23, 2–3, 231–264.

Norde, M. (2009). *Degrammaticalization.* Oxford: Oxford University Press.

Ohala J. J. (1980). The acoustic origin of the smile. *Journal of the Acoustical Society of America*, 68, S33.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (ed.), *Historical Linguistics: Problems and Perspectives* (pp. 237–78). New York: Longman.

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, and J. J. Ohala (eds), *Sound Symbolism* (pp. 325–47). Cambridge: Cambridge University Press.

Ohala, J. J. (2003). Phonetics and historical phonology. In Joseph and Janda (2003), pp. 669–86.

Penny, R. (1991). *A History of the Spanish Language.* Cambridge: Cambridge University Press.

Philippaki-Warburton, I. (1994). The subjunctive mood and the syntactic status of the particle *na* in Modern Greek. *Folia Linguistica*, 28, 297–328.

Rask, R. (1818). *Undersögelse om det gamle Norske eller Islandske Sprogs Oprindelse.* Copenhagen: Gyldendal.

Ringe, D. and Eska, J. F. (2013). *Historical Linguistics: Toward a Twenty-First Century Reintegration.* Cambridge: Cambridge University Press.

Romaine, S. and Lange, D. (1991). The use of *like* as a marker of reported speech and thought: a case of grammaticalization in progress. *American Speech*, 66, 227–79.

Salmons, J. C. and B. D. Joseph. (1998). Introduction. In J. C. Salmons and B. D. Joseph (eds.), *Nostratic: Sifting the Evidence* (pp. 1–9). Current Issues in Linguistic Theory, 142. Amsterdam: John Benjamins.

Schourup, L. (1985 [1982]). *Common Discourse Particles in English Conversation.* Outstanding Dissertations in Linguistics. New York: Garland. (Revised version of 1982 PhD dissertation, Ohio State University.)

Shoaf, R. A. (ed.) (1989). *Geoffrey Chaucer: Troilus and Criseyde.* East Lansing, MI: Colleagues Press.

Teodorsson, S.-T. (1979). Phonological variation in Classical Attic and the development of koine. *Glotta*, 57, 61–75.

Thomason, S. G. (2003). Contact as a source of language change. In Joseph and Janda (2003).

Thomason, S. G. and Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics.* Berkeley, CA: University of California Press.

Trask, R. L. (1996). *Historical Linguistics.* London: Arnold. (1st edn; 2nd edn as Millar 2013.)

Traugott, E. C. and Dasher, R. B. (2002). *Regularity in Semantic Change.* Cambridge: Cambridge University Press.

Verner, K. (1877). Eine Ausnahme der ersten Lautverschiebung. *Zeitschrift für vergleichende Sprachforschung*, 23, 97–130. (Translated in W. Lehmann (ed.), *A Reader in 19th Century Historical Indo-European Linguistics.* Bloomington: Indiana University Press.)

Watkins, C. W. |(1962). *Indo-European Origins of the Celtic Verb.* Dublin: Institute for Advanced Studies.

Weinreich, U., Labov, W., and Herzog, M. I. (1968). Empirical foundations for a theory of language change. In W. Lehmann and Y. Malkiel (eds), *Directions for Historical Linguistics* (pp. 95–195). Austin, TX: University of Texas Press.

Winters, M. (1995). The nature of the so-called analogical processes [translation of Kurylowicz 1947]. *Diachronica: International Journal for Historical Linguistics*, 12, 1, 113–45.

Part IV Languages and the Mind

16 Neurolinguistics*

DAVID CAPLAN

The field of neurolinguistics has come to consist of two related areas of study – language disorders (which is sometimes called "aphasiology") and the relationship between language and the brain. Aphasiology has made important discoveries about what goes wrong with language after brain damage, some of which have implications for what language is and how it is processed. Aphasiology is closely linked to the study of the relationship between language and the brain. The combined study of patients' language deficits and neurological lesions provides evidence about the location and type of brain damage that affects language and therefore about the brain areas that are necessary to perform particular language functions. This approach been considerably refined as more detailed descriptions of language disorders have been produced that draw on linguistics and psychology and as advances in neuroimaging allow for much more precise and complete characterization of lesions than previously possible.

In recent years, the traditional deficit–lesion correlational approach to brain organization for language has been complemented by studies of brain function in normal subjects when they perform tasks that involve language. "Functional neuroimaging" uses positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), which primarily measure regional cerebral blood flow (rCBF). Measurements can also be made on the scalp of electrophysiological potentials ("event related potentials" – ERPs) and small electromagnetic potentials (magnetoencephalography – MEG) to record neural activity that arises in relationship to language functions. The development of these techniques has made the present period one of the most exciting in the history of neurolinguistics. In this chapter, I shall present selected results of work in this rapidly evolving field.

1 Aphasiology

1.1 A very brief history of aphasiology

Paul Broca's presentation to the Anthropological Society of Paris of the now famous case of Lebourgne (Broca 1861) started and defined both aphasiology and neurolinguistics. Broca's case was a 57-year-old man who had come into the hospital 21 years earlier unable to speak. He could only utter the syllable "tan." He appeared able to understand what people said to him, and he could

* This work was supported by a grant from the National Institute for Neurological Disease and Stroke (DC00942).

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

express his needs and emotions through gesture and by changing the tone of his voice. Over the years he developed a right-sided weakness, and eventually became bedridden. He died of an infection that began in bed sores in his right leg, which had become ulcerated and infected as a result of having been lain on and improperly attended to. For the last ten years of his life, his language capacities remained static – at least according to the information that Broca had and that he relayed to the Anthropological Society.

Broca's analysis of the language disorder in "Tan" was that he had lost the "faculty of articulate speech" while retaining the faculty for language comprehension and those for production and recognition of meaningful gestures. With this analysis, Broca defined the approach to aphasia that dominated medical and scientific thinking for decades – to think of language disorders as disorders of the ability to use language in the usual tasks of speaking, understanding, reading, and writing. This approach focused attention on the common tasks by which linguistically based communication takes place. This is a domain of human linguistic functioning that, though complicated, is much more tractable and much better understood than domains such as verbally mediated planning, self-expression, verbal creativity, etc. From the neurolinguistic point of view, these basic functions of speech, comprehension, reading, and writing are related to a relatively small part of the brain. Thus, this focus allowed researchers to make some headway into the question of how the brain is related to language functions.

From the point of view of aphasiology, the major contribution of the work that followed and built upon Broca's discovery was the description of the "classic" clinically defined aphasic syndromes, which still are commonly referred to by clinicians who diagnose and treat language disorders. These "aphasic syndromes" are shown in Table 16.1. "Broca's aphasia" is a severe expressive language disturbance reducing the fluency of speech without an equally severe disturbance of auditory comprehension. "Wernicke's aphasia" consists of the combination of fluent speech with erroneous choices of the sounds of words (phonemic paraphasias) and an auditory comprehension disturbance. "Pure word deafness" is the relatively pure case of an auditory receptive disorder in which the patient does not recognize spoken words, so spontaneous speech is normal but comprehension and repetition are disturbed. "Anarthria," "dysarthria," and "apraxia of speech" are output speech disorders in which both repetition and spontaneous speech are misarticulated but comprehension is preserved. In "transcortical motor aphasia," spontaneous speech is reduced but repetition is intact; in "transcortical sensory aphasia," a comprehension disturbance exists without a disturbance of repetition. A disturbance in spontaneous speech and repetition without a disturbance in auditory comprehension is termed "conduction aphasia." All of these syndromes were claimed to have been discovered in relatively pure form by researchers by 1885 (Lichtheim 1885). This classification was revived by Normal Geschwind in the 1960s (Geschwind 1965).

Table 16.1 The classic aphasic syndromes.

Syndrome	Clinical Manifestations	Hypothetical Deficit	Classical Lesion Location
Broca's aphasia	Major disturbance in speech production with sparse, halting speech, often misarticulated, frequently missing function words and bound morphemes	Disturbances in the speech planning and production mechanisms	Primarily posterior aspects of the 3rd frontal convolution and adjacent inferior aspects of the precentral gyrus

(continued)

Syndrome	Clinical Manifestations	Hypothetical Deficit	Classical Lesion Location
Wernicke's aphasia	Major disturbance in auditory comprehension; fluent speech with disturbances of the sounds and structures of words (phonemic, morphological, and semantic paraphasias)	Disturbances of the permanent representations of the sound structures of words	Posterior half of the first temporal gyrus and possibly adjacent cortex
Anomic aphasia	Disturbance in the production of single words, most marked for common nouns with variable comprehension problems	Disturbances of the concepts and/or the sound patterns of words	Inferior parietal lobe or connections between parietal lobe and temporal lobe
Global aphasia	Major disturbance in all language functions	Disruption of all language processing components	Large portion of the peri sylvian association cortex
Conduction aphasia	Disturbance of repetition and spontaneous speech (phonemic paraphasias)	Disconnection between the sound patterns of words and the speech production mechanism	Lesion in the arcuate fasciculus and/or cortico-cortical connections between temporal and frontal lobes
Transcortical aphasia	Disturbance of spontaneous speech similar to Broca's aphasia with relatively preserved repetition	Disconnection between conceptual representations of words and sentences and the motor speech production system	White matter tracts deep to Broca's area
Transcortical sensory aphasia	Disturbance in single word comprehension with relatively intact repetition	Disturbance in activation of word meanings despite normal recognition of auditorily presented words	White matter tracts con necting parietal lobe to temporal lobe or in portions of inferior parietal lobe
Isolation of the language zone	Disturbance of both spontaneous speech (similar to Broca's aphasia) and comprehension, with some preservation of repetition	Disconnection between concepts and both representations of word sounds and the speech production mechanism	Cortex just outside the perisylvian association cortex

These syndromes mainly reflect the relative ability of patients to perform language tasks (speaking, comprehension, etc.), not the integrity of specific components of the language processing system. For instance, the speech production problem seen in Broca's aphasia can consist of one or more of a large number of impairments – disorders affecting articulation such as dysarthria or apraxia of speech, disorders affecting sentence form such as agrammatism, etc. Patients with Wernicke's aphasia can have deficits affecting either the sounds of words or their meanings

or both, as well as any number of other language processing deficits. There are at least two major deficits that underlie "conduction aphasia," one affecting word production and one affecting verbal short-term memory. At the same time as patients with the same syndrome can have different deficits, identical deficits occur in different syndromes. For instance, certain types of naming problems can occur in any aphasic syndrome. For these reasons, knowing that a patient has a particular aphasic syndrome does not tell us exactly what is wrong with his/her language, and is of limited use in understanding the detailed nature of language impairments, the areas of the brain needed for specific language operations, and even for clinical purposes such as planning therapy (Schwartz 1984). For that, one has to turn to more detailed studies.

1.2 Modern aphasiology

The contemporary approach to language disorders sees human language as a code that links a set of linguistic forms to a number of aspects of meaning. These forms are activated in the usual tasks of language use – speaking, auditory comprehension, reading, and writing – by different "components" of a "language processing system." For instance, a component of the language processing system might accept as input the semantic representation (meaning) activated by the presentation of a picture and produce as output a representation of the sound pattern of the word that corresponds to that meaning.

At least four levels of language representation are clearly identified in contemporary models of the language processing system: the *lexical level*, the *morphological level*, the *sentential level*, and the *discourse level*. Details regarding the structure of language are discussed in Chapters 9–14. These levels of the language code are all activated in parallel in the acts of speaking, understanding, reading, and writing. Speaking provides an example. A speaker selects words to convey concepts on the basis of his/her knowledge of a subject and also as a function of what words s/he thinks the listener will understand. S/he selects the sounds of each word and their order. S/he puts the words into the grammatical structures needed to convey the relationships between their meanings and finds the right morphological form for each word in each grammatical position. S/he selects intonational contours that express whether a sentence is a statement, a question, a command, or has some other illocutionary force. Intonational contours and syntactic structures are also selected to express the focus of the discourse. The speaker sends commands to the muscles of the chest, the diaphragm, the larynx, and the oral and nasal tracts to integrate the movements of all these regions in a way that produces exactly the right sounds for each word with the right degree of stress. All these operations go on unconsciously, at a speed of about 120 words per minute, or 2 words per second, or roughly 1 sound (phoneme) every 100 milliseconds or so. A speaker is not only fast at doing these remarkable computations and retrieval operations; s/he is also accurate. Estimates of the number of errors that a speaker makes are in the range of one semantic mistake every million words and a comparable number of sound-based errors (see Levelt 1989, for an extensive review of these processes).

Speaking is a remarkable act, but no more so than understanding spoken speech (where the words, structure, and meaning of an utterance are extracted from the acoustic signal), or reading or writing. All these functions are the result of the integrated activity of dozens of highly specific operations acting in parallel and in critically ordered sequences. More detailed discussion of language processing can be found in Chapter 17.

Modern aphasiology tries to understand disturbances of speaking, comprehension, reading, and writing as disruptions of specific language processing components. It is impossible to present a review of all the disturbances that have been described that affect each of the components of the language processing system in each of the tasks of speaking, comprehending, reading, and writing. I shall instead sample from two areas of language – disorders affecting patients' abilities to understand the meanings of words and disorders affecting their abilities to understand sentences.

1.3 Disturbances of word meanings

Most recent research on disturbances of word meanings in brain-damaged patients has focused on words that refer to objects. The meanings of these words are thought to be stored in a specialized memory store, called "semantic memory" (Tulving 1972, 1983). Though semantic memory is also thought to house representations of entities other than objects (such as events, for instance), the concepts that correspond to objects have been the subject of the most extensive thought and investigation in philosophy and psychology. The representations of objects in semantic memory have traditionally been thought of as sets of features that list the properties that are necessary and sufficient for an item to fall into a given category, such as that for an entity to be a bird it has to have feathers and to fly. It is now appreciated that most concepts are not easily described in these terms, and that categorization must depend on probabilistic criteria (see Smith and Medin 1981, for discussion). For instance, penguins and ostriches are birds but do not fly. The concepts corresponding to concrete objects may be represented in both a "verbal" and a "visual" semantic memory system (Miller and Johnson-Laird 1976; Paivio 1971), though the evidence for this is controversial (Snodgrass 1984). The verbal semantic system may encode relations of one entity to another and functional properties of each item, such as the fact that some birds fly south in the winter. The visual semantic system may encode physical properties of an item that allow it to be recognized, such as the typical silhouette of a bird.

Disturbances of word meanings thus can be seen as problems affecting the ability of a patient to appreciate how an item is classified or what its necessary and typical features are. Such disturbances would be expected to affect a patient's ability to identify a word or a picture as an example of a specific concept, and therefore to cause poor performance on word–picture matching and naming tasks (Warrington 1975). The co-occurrence of deficits in word–picture matching and naming is not an adequate basis for diagnosing a problem affecting word meaning, however, because a patient may have separate disturbances that affect word recognition and production independently (Howard and Orchard-Lisle 1984). Co-occurring deficits in naming and word–picture matching are more likely to result from a disturbance affecting concepts when the patient makes many semantic errors in providing words to pictures and definitions, s/he has trouble with word–picture matching with semantic but not phonological foils, s/he fails on categorization tasks with pictures, and when the same words are affected in production and comprehension tasks (Hillis et al. 1990).

It has been argued that brain damage may affect either the storage or the retrieval of word meanings. Shallice (1988a, 1988b), Warrington and Shallice (1979), and Warrington and McCarthy (1987) have suggested that there are five hallmarks of the loss of items in semantic memory: consistent production of semantic errors on particular items across different inputs (pictures, written words, spoken words); relative preservation of superordinate information as opposed to information about an item's features; relative preservation of information about higher frequency items; no improvement of performance by priming and cueing; and no effect of the rate at which a task is performed upon performance. These researchers have also suggested that disorders of retrieval of items from semantic memory is characterized by the opposite effects of these variables on performance. These authors have described patients with semantic impairments, whose impairments they say are of one or the other of these types. The criteria for distinguishing between storage and retrieval impairments remain controversial, however (Caramazza et al. 1990; see Caplan 1992: ch. 4, for discussion).

Disorders affecting processing of semantic representations for objects may be specific to certain types of inputs. Warrington (1975) first noted a discrepancy between comprehension of words and pictures in two dementing patients. Bub et al. (1988) have analyzed a patient, M. P., who showed very poor comprehension of written and spoken words but quite good comprehension of pictures. These impairments have been taken as reflections of disturbances of "verbal" and "visual" semantic systems, though this interpretation is debated (Riddoch and Humphreys 1987).

Semantic disturbances may also be category-specific. Several authors have reported a selective semantic impairment of concepts related to living things and foods compared to human-made objects (Sartori and Job 1988; Silveri and Gainotti 1988; Warrington and Shallice 1984). The opposite pattern has also been found (Warrington and McCarthy 1983, 1987). Selective preservation and disruption of abstract versus concrete concepts, and of nominal versus verbal concepts, have also been reported (Miceli et al. 1984; Schwartz, Marin, and Saffran, 1979; Warrington 1981a, 1981b; Zingeser and Berndt 1988). The reasons for category-specific deficits have been a topic of interest. Some authors argue that they reflect the properties of different categories, with biological categories being distinguished from one another more on the basis of the physical features of their members than their functions, and vice versa for manmade artifacts (Cree and McRae, 2003; Crutch and Warrington, 2003; Humphreys and Forde, 2001). Others have suggested that category-specific semantic deficits occur because the conceptual knowledge corresponding to objects with similar properties is stored in adjacent neural areas or because damage to one property spreads to highly associated properties (Tyler and Moss 2001; Tyler et al. 2000; Devlin et al. 1998; Garrard et al. 2001; Gonnerman et al. 1997; McRae and Cree 2002; Vinson et al. 2003). Another possibility is that certain major categories – faces, foods, animals – are innately encoded neurologically (Caramazza and Shelton 1998; Shelton, Fouch, and Caramazza 1998).

Aphasic disorders may affect conscious and unconscious semantic processing differently. Milberg and his colleagues (Milberg and Blumstein 1981; Blumstein, Milberg, and Shrier 1982) have described patients who cannot match words to pictures or name objects, but who show evidence of understanding words unconsciously. The evidence that these patients process meaning unconsciously is that they show "semantic priming effects." These effects consist of responding more quickly to a word when it has been preceded by a semantically related word in a task that does not require the subject to process the word's meaning. For instance, if a subject is shown sequences of letters on a computer screen and asked to press a button to indicate whether a sequence is a word or not (a "lexical decision" task), s/he will respond faster to the letter string DOCTOR when it follows the string NURSE than when it follows the string TABLE. Milberg and his colleagues have found these priming effects in some patients who do not show evidence of understanding words when the task requires conscious processing of a word's meaning, suggesting that they are able to appreciate the meanings of words unconsciously but not process meaning in conscious, controlled tasks like word–picture matching. Conversely, Swinney, Zurif, and Nicol (1989) have shown that some patients who appear to understand words well may have abnormalities in tasks that are sensitive to unconscious processing of the meanings of words. These researchers reported that four aphasic patients who were good at word–picture matching performed abnormally in a priming task with ambiguous words (like *bank*). Unlike neurologically normal individuals, who showed priming for both senses of an ambiguous word, the four patients studied by Swinney and his colleagues only showed priming for the most frequent sense of ambiguous words. This indicates that their unconscious processing of word meanings was abnormal, even if it did not affect their performance on untimed, controlled, conscious tasks such as word–picture matching.

1.4 Disorders of sentence comprehension

When a subject understands a sentence, s/he combines the meanings of the words into a propositional content in accordance with the syntactic structure of the sentence. There are many reasons why an aphasia patient might fail to carry out the operations that are needed to arrive at propositional content. Disturbances affecting comprehension of simple and morphologically complex words affect comprehension at the sentence level. In addition, there are disturbances affecting patients' abilities to understand aspects of propositional meaning despite good single word comprehension.

The largest amount of work in the area of disturbances of sentence comprehension has gone into the investigation of patients whose use of syntactic structures to assign meaning is not

normal. Caramazza and Zurif (1976) were the first researchers to show that some patients have selective impairments of this ability. These researchers described patients who could match "semantically irreversible" sentences such as *"The apple the boy is eating is red"* to one of two pictures but not "semantically reversible" sentences such as *"The girl the boy is chasing is tall."* The difference between the two types of sentences resides in the fact that a listener can understand a sentence such as *"The apple the boy is eating is red"* because boys are animate and can eat, and apples are inanimate and can be eaten, whereas understanding *"The girl the boy is chasing is tall"* requires assigning the syntactic structure of the sentence since both boys and girls are capable of chasing one another. Caramazza and Zurif concluded that their patients could not assign or use syntactic structure for this purpose.

Disorders of syntactic comprehension have since been examined in considerable detail (Caplan and Futter 1986; Caplan and Hildebrandt 1988; Caplan, Waters, and Hildebrandt 1997; Caplan, Waters, DeDe et al. 2007; Hildebrandt, Caplan, and Evans 1987, Linebarger, Schwartz, and Saffran 1983; Schwartz, Saffran, and Marin 1980; Tyler 1985). Patients may have very selective disturbances affecting the use of particular syntactic structures or elements to determine the meaning of a sentence. For instance, two patients we studied showed a double dissociation in their abilities to understand sentences with reflexive elements (*himself*) and pronouns (*him*) (Caplan and Hildebrandt 1988). Some patients can understand very simple syntactic forms, such as active sentences (*The man hugged the woman*), but not more complex forms, such as passive sentences (*The woman was hugged by the man*) (Caplan, Baker, and Dehaut 1985). Many other studies have shown apparently isolated disorders affecting syntactically characterized elements of language ranging from types of verbs through the meanings of quantifiers.

Some patients appear to be able to assign syntactic structure, but not to use it to determined sentence meaning. Linebarger and her colleagues (Linebarger, Schwartz, and Saffran 1983; Linebarger 1990) have reported that some patients who have syntactic comprehension problems (who cannot match reversible sentences to pictures, for instance) can make judgments as to whether or not a sentence is grammatical. For instance, some patients can indicate that the utterance *"The woman was watched the man"* is ill formed and the utterance *"The woman was watched by the man"* is acceptable, despite not being able to match sentences such as *"The woman was watched by the man"* to one of two pictures. These researchers have interpreted these results as an indication that some patients can construct syntactic structures but not use them to determine propositional meaning (a so-called "mapping" problem – Schwartz, Linebarger, and Saffran 1985). Other patients have virtually no ability to use syntactic structure at all.

Many patients with such disorders use strategies such as assigning the thematic role of agent to a noun immediately before a verb to understand semantically reversible sentences, leading to systematic errors in comprehension of sentences such as *"The boy who pushed the girl kissed the baby."* Most patients with these disorders appear to rely upon inferences based upon their knowledge of the real world and their ability to understand some words in a sentence, as seems to have been the case with the original patients described by Caramazza and Zurif (1976).

The evidence that supports or disconfirms these models is almost exclusively derived from end-of-sentence comprehension performance. The major advance in the past three decades of study of aphasic deficits of syntactically based comprehension is the application of online observational methods to the question of what underlies these abnormal performances. The evidence is sparse, but a common thread runs through it – online processing of syntactic structure often appears to be much more intact in aphasic patients than was thought on the basis of end-of-sentence task performance. Tyler (1985) reported that an agrammatic aphasic, DE, whose end-of sentence anomaly judgments were much less accurate than those of neurologically normal individuals, showed several features in a word monitoring task that suggest intact online parsing. Swinney and Zurif (1995) reported a series of lexical priming studies in which they found lexical semantic priming for written words immediately after the presentation of the head noun of a relative clause and again at a later point at which the word would have been reaccessed (the verb in the relative clause) but not at an intermediate point during

the auditory presentation of a sentence, and interpreted this pattern as evidence for reactivation of the antecedent of the trace at that point. They reported that eight anomic patients who performed poorly on end-of-sentence comprehension measures showed this pattern. Blumstein et al. (1998) reported this pattern in several Broca's aphasics. These results provide evidence for the integrity of some syntactic and interpretive operations in both the major clinical groups of aphasic patients.

An important advance has been the separation of online observations made in trials on which end-of-sentence responses are correct and erroneous. Caplan, Waters, DeDe et al. (2007) found that, when patients whose accuracy on some sentence types was at chance made correct responses, their self-paced listening times (corrected for spoken word duration and word frequency) in auditory moving windows presentation and sentence-picture and grammaticality judgment tasks showed normal effects of syntactic structure, but that self-paced listening times were abnormal at critical points in complex sentences when patients made errors. Caplan and Waters (2003) found that, for cleft structures (e.g., *It's you who left*), poor comprehending patients had longer residual listening times for the embedded verb in cleft object sentences than for the embedded verb or noun phrase in cleft subject sentences to which they responded correctly. They had longer residual listening times for the embedded verb in cleft object sentences to which they responded erroneously than for the embedded verb in cleft object sentences to which they responded correctly. These residual listening times in cleft object sentences suggests these low comprehending patients attempted to assign the structure and meaning of these sentences online. When they succeeded, their residual listening times showed the normal pattern; failures were associated with continued, unsuccessful efforts to parse and interpret the sentences, seen at the most demanding phrase of the sentence.

Other patients show normal signs of processing at the points of increased demand, but problems after that point. Dickey and Thompson (2004) reported that 12 Broca's aphasics showed normal eye fixations on pictures of words designated in a spoken sentence at the embedded verb in object extracted structures (*wh* -questions), despite making many more errors than controls; errors were associated with increased looks to the subject noun phrase later in the presentation of the sentence. Thompson and Choy (2009) reported similar findings in eight agrammatic patients at points of presentation of pronouns and reflexives. These authors suggested that their patients initially processed the sentences normally, to the point of initially understanding the thematic (semantic) roles in object extracted structures and finding the antecedents of reflexives and personal pronouns in the same way as normals. In their view, errors were due to patients' initial understanding being overridden at a later point in processing by an alternative interpretation of the sentences. The difference between aphasics and normals, on this view, is not that aphasics' initial comprehension processes are disturbed, but that aphasics are less capable than normals of determining that a syntactically derived meaning, as opposed to a meaning derived in some other fashion, is correct. This deficit would be a failure of some sort of control, or perhaps of a labeling process, not of assigning syntactic structure or using it to determine sentence meaning per se.

A well-known set of results are the reports by Swinney and his colleagues (Swinney and Zurif 1995; Zurif et al. 1993) that some Broca's aphasics did not show the lexical semantic priming pattern described above that they interpreted as evidence for online reactivation of the head noun of a relative clause; the absence of priming at the embedded verb was interpreted as indicating an online impairment activating the antecedent of a trace, consistent with the trace deletion hypothesis (Grodzinsky 1990). It has become clear, however, that those studies cannot be unequivocally interpreted in this fashion, for two reasons. First, the authors did not separate correct and erroneous trials. Second, the Broca's aphasics tested in the Swinney/Zurif studies showed no priming effects at any position, not just at the verb of the relative clause, and were not tested for word-to-word priming effects in isolation; the absence of any priming effects may thus have been due to a failure of these patients to show lexical priming, as has been documented for Broca's aphasics by Milberg and Blumstein (1981; Blumstein, Milberg, and Shrier 1982). Follow-up studies in other Broca's aphasics (Love, Swinney, and Zurif 2001; Love et al. 2008) showed delayed priming after both the initial

presentation of the head noun and the relative clause verb, leading Love et al. to conclude that the patients experienced a delay in activation of lexical semantic or conceptual representations that interfered with parsing and interpretation. This conclusion was supported by the finding that the patients' comprehension improved significantly when the sentences in which the relative clause modified the object were presented at a slower speed (see also Friederici and Kilborn 1989; Haarmann and Kolk 1991, for models that involve slowing of lexical activation or syntactic processing).

All the studies reviewed above have interpreted both accuracy and online measures as reflections of online syntactic processing and comprehension. However, the picture appears to be more complicated. Caplan, Waters, DeDe et al. (2007) reported that patients' self-paced listening times were longer for the same sentences in a sentence–picture matching than in a grammaticality judgment task, suggesting that these responses partly reflect the time it takes to begin to accomplish the task while patients are self-pacing themselves through a sentence word-by-word. Accuracy results show that many deficits affect certain structures in only one task (Linebarger, Schwartz, and Saffran 1983; Linebarger 1995; Cupples and Inglis, 1993; Caplan, DeDe, and Michaud 2006; Caplan, Waters, DeDe et al. 2007). How deficits affect the interactions between assigning sentence structure, assigning sentence meaning, and performing a task remains to be explained.

I will end this section with a comment on the relation between deficits of short-term (working) memory and syntactically based comprehension. Disturbance of short-term memory can produce sentence comprehension impairments, but the connection between the two disorders is more complex than one might suspect. Martin and He (2004) have related short-term memory impairments to difficulties in understanding parts of sentences that consist of lists of words, such as *big*, *noisy*, and *aggressive*, in the sentence *The neighbor's dog was big, noisy, and aggressive*. However, many case studies show that patients with short-term memory impairments can have excellent syntactic processing abilities (Caplan and Waters 1990; McCarthy and Warrington 1984; Butterworth, Campbell, and Howard 1986; Waters, Caplan, and Hildebrandt 1991). Short-term memory impairments may induce problems remembering the meaning of a sentence, not understanding that meaning in the first place (see Caplan and Waters 1999, for review).

1.5 Comments on modern aphasiology

This brief overview merely conveys some of the results of recent research into disorders affecting word meanings and sentence comprehension. Even this brief survey indicates that these disturbances are very complex, and vary in different patients. It shows that studies of brain-damaged patients can suggest features of how language is structured, such as the suggestion that word meanings are organized into semantic categories, or that the assignment of syntactic structure is partially independent of the use of that structure to determine sentence meaning. The review indicates that modern aphasiology provides suggestions regarding representations and operations that need to be related to the brain: verbal semantic representations, visual semantic representations, category-specific semantic representations, specific syntactic operations, etc. We now turn to this second aspect of neurolinguistics.

2 Language and the Brain

2.1 The overall organization of the brain for language

The brain is perhaps the most highly differentiated organ in the body. It consists of a large number of regions, each of which contributes to sensation, motor function, thought, emotion, and other functions in special ways. Only a relatively small part of the brain is devoted to language (Figure 16.1). This part primarily lies in the cerebral cortex – a thin strip of neural cells and supporting tissue along the outermost edge of the brain – and consists of the association cortex in

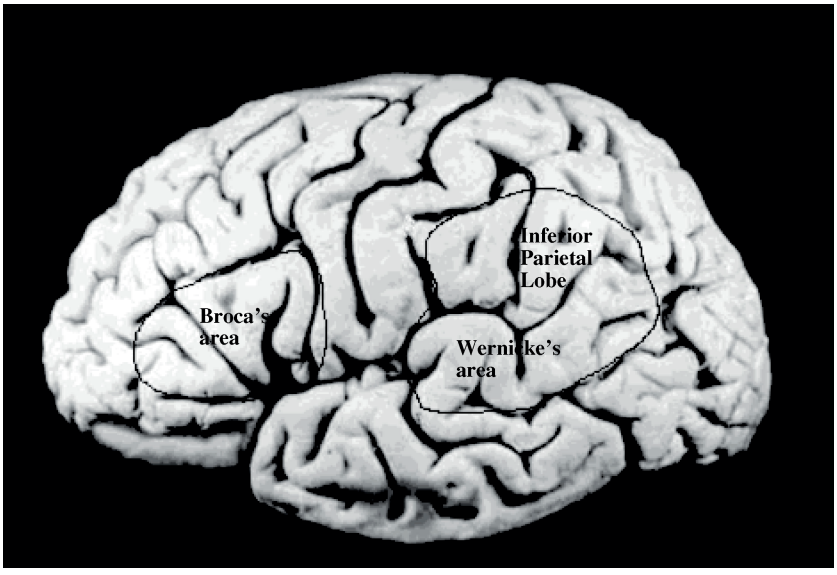


Figure 16.1 Depiction of the left lateral surface of the human brain showing the perisylvian association cortex related to language.

the region of the sylvian fissure (Luria 1970; Russell and Esper 1961; Basso et al. 1985; Weisenberg and McBride 1935; Brown 1972; Pick 1973). This region includes the pars triangularis and the pars opercularis of the third frontal convolution (Broca's area), the association cortex in the opercular area of the pre- and post-central gyri, the supramarginal and angular gyri of the parietal lobe, the first temporal gyrus from the supramarginal gyrus to a point lateral to Heschl's gyrus (Wernicke's area), and possibly a portion of the adjacent second temporal gyrus. On the basis of present knowledge, this is the primary cortical area that can be confidently thought to subserve language functions in the aural-auditory modality of use. Other areas may be involved in language in other modalities. There is strong evidence for a specialization of cortex in the inferior temporal lobe for letter and/or word recognition (Dehaene and Cohen 2011) and signed language appears to involve cortex somewhat more distant from the perisylvian fissures and closer to motor and sensory cortex associated with planning and executing hand motion. Yet other cortical areas play roles in control of language functions. For instance, the supplementary motor area appears to be to be important in initiating vocalization (Masdeu, Schoene, and Funkenstein 1978).

The cortex does not exist or function alone; it is connected by large white matter tracts to subcortical nuclei in the basal ganglia, the thalamus, and the cerebellum. These structures are part of a "neural system" that processes language. Lesions of both the white matter tracts that connect parts of the language cortex to these subcortical regions and to one another produce language impairments (Naeser et al. 1982; Cappa et al. 1983; Damasio et al. 1982; Mohr, Watters, and Duncan 1975). However, the exact function of these structures is not completely understood. The white matter tracts do not compute representations but only transfer the results of computations from one set of neurons to another. The basal ganglia, thalamus, and cerebellum may actually store and compute linguistic representations, and some researchers have suggested that these subcortical grey matter structures play roles in language processing (Damasio et al. 1982; Mohr, Watters, and Duncan 1975; Crosson 1985). However, an alternative possibility is that the language disorders that follow subcortical lesions result from altered physiological activity in the overlying cortex, not disorders of the subcortical structures themselves.

The availability of patients with focal strokes that are visible only subcortically on CT or MR scans, in whom metabolic scanning is used to assess lesion site and size in both cortical and subcortical structures, provides an opportunity to investigate the role that both cortical and subcortical structures play in language. Across all the published cases of subcortical aphasia, there is a 1:1 correspondence between the presence or absence of cortical hypometabolism or hypoperfusion (lowered glucose utilization and lowered blood flow) and the presence or absence of aphasic impairments in patients with strokes that are visible only subcortically on CT scans (Perani et al. 1987; Olsen, Bruhn, and Oberg 1986). Moreover, studies correlating the degree of hypometabolism measured cortically and subcortically with the degree of language impairment indicate a much higher correlation of language impairments in aphasic patients with the indices of cortical hypometabolism (Metter et al. 1988; Metter et al. 1983; Kempler et al. 1988; Metter et al. 1987). There is not a single published case in which an aphasia has been documented in a patient in whom metabolic scanning, blood flow studies, and physiological measures have all shown normally functioning perisylvian association cortex. The conclusion that is suggested by these results is that subcortical structures are not themselves responsible for language processing but serve only to activate the cortically based language processing system and to transfer the results of psycholinguistic computations from one part of the perisylvian association cortex to another. The simplest model consistent with available data is that language processing is carried out only in the perisylvian association cortex. However, we have observed aphasic patients with what appear to be purely subcortical lesions that spare the cortex (the integrity of the cortex was ascertained by structural and perfusion MR) and this conclusion may be modified as more information accumulates.

A well-attested feature of neural organization for language is lateralization – the fact that language processing relies upon one hemisphere more than another in most normal individuals. In about 98 percent of strong right-handers from right-handed families, the left perisylvian association cortex accomplishes most, if not all, language processing functions (Luria 1970; Milner, Branch, and Rasmussen 1964; Milner 1974). In individuals with other handedness profiles (Geschwind and Galaburda 1987, 1985; Annett 1985), language functions are far more likely to involve the corresponding regions of the right hemisphere (Luria 1970; Russell and Esper 1961; Goodglass and Quadfasel 1954), with different likelihoods of right and left hemispheric involvement in language functions in different subgroups within this population (Subirana 1964). The data on differential lateralization as a function of sex are controversial (McGlone 1980).

A potentially important point is that many aphasic syndromes that follow either left or right hemisphere lesions in subjects who are not right-handed are often mild. Their occurrence suggests that many individual language processing components can be located in either hemisphere. Whether these language processing components are located in a given hemisphere in isolation from others can only be resolved by studies that establish whether the remaining intact language components are based in the intact portions of the lesioned hemisphere or in the undamaged hemisphere of patients with mild aphasias. In some cases (Kinsbourne 1971), intracarotid amytal injections (Wada studies) indicate that the latter appears to be the case. This would suggest separate control of lateralization for individual language processing components, but very few data are available on this point. Functional neuroimaging after stroke has shown variable patterns of hemispheric involvement. Over time, both intact ipsilesional (on the same side as the lesion) and homologous contralesional (on the opposite side from the lesion) areas have shown changes in neurovascular activity as language improves, with some evidence of greater ipsilesional activity associated with recovery as time passes (Cloutman et al. 2009). These data suggest that there is a considerable degree of plasticity for language even in adult brain.

Though not as critical to language functioning as the dominant hemisphere, the nondominant hemisphere is involved in many language operations. Evidence from the effects of lesions and split brain studies, as well as experiments using presentation of stimuli to one or the other hemisphere in normal subjects, indicate that the nondominant hemisphere understands many words, especially concrete nouns (Gazzaniga 1983; Chiarello et al. 1990), and suggest that it is involved in

other aspects of language processing as well. For instance, our studies have shown effects of right hemisphere stroke upon syntactic processing, although these are much more mild than those found after left hemisphere strokes (Caplan, Hildebrandt, and Makris 1996), and this result has its counterpart in at least one study that showed increases in blood flow in the right hemisphere that were associated with syntactic processing (Just et al. 1996; Ben Shachar, Palti, and Grodzinsky 2004). Some language operations may be carried out primarily in the right hemisphere. The best candidates for these operations are ones that pertain to processing the discourse level of language (especially revising inferences that have been made), interpreting nonliteral language such as metaphor, and appreciating the tone of a discourse as is manifest in, for instance, humor (Brownell and Gardner 1988; Joanette and Brownell 1990; Molloy, Brownell, and Gardner 1990; Roman et al. 1987). Some scientists have developed models of the sorts of processing that the right hemisphere carries out. For instance, Beeman (1998) has suggested that the right hemisphere codes information in a coarse way compared to the left. This and other suggestions provide the bases for ongoing research programs into the nature of language processing in the right hemisphere.

2.2 *The organization of the perisylvian association cortex for language*

Historically, theories of the relationship of parts of the perisylvian association cortex to components of the language processing system have included distributed ("holist") and localizationist models. Contemporary studies have strongly supported the latter, and representative work will be selectively reviewed.

Evidence for localization of language processing components comes from the finding that selective language deficits occur in patients with perisylvian lesions, often in complementary functional spheres. For instance, patients with agrammatism have trouble producing the grammatical function words of language whereas patients with anomia have trouble producing common nouns. The existence of these two disorders indicates that the tissue involved in producing function words is not involved in producing common nouns in the first set of patients, and vice versa in the second set.

The first localizationist theories were based on clinical observations in the mid and late nineteenth century, and were linked to the classical aphasic syndromes described above. Patients with lesions in the left inferior frontal lobe were found to have hesitant and poorly articulated speech and patients with lesions in the superior temporal lobe to have disturbances of comprehension and fluent speech with sound and word substitutions. These correlations led to the theory that motor speech planning required Broca's area (Brodman areas 44 and 45) and that the representations of the sounds of words (the lexicon for both speech and spoken word recognition) was localized in Wernicke's area (Brodman area 22). This model emphasized the connectivity and proximity of these regions to primary motor and auditory cortex. Geschwind (1965) added the hypothesis that word meaning was localized in the inferior parietal lobe (Brodman areas 39 and 40), noting that the inferior parietal lobe is an area of multimodal association cortex to which fibers from the unimodal association cortex related to audition, vision, and somasthesis project, and that word meanings consist of associations between sounds and properties of objects. These models are associated with the deficit analyses proposed in the classical aphasic syndromes, and are correspondingly limited. They largely deal with words to the relative exclusion of other levels of the language code, and some functions that are localized are entire language tasks, such as speech production. They also suffer from empirical problems. The correlations between syndromes and lesions are found primarily in the chronic stage of recovery from stroke, not in acute or subacute phases of stroke or in other diseases, and even in chronic strokes a very large number of lesions do not occur where the model predicts (Kolb and Whishaw 2003). Modern work has disconfirmed some basic features of the model, such as Geschwind's proposal that word meanings are supported by the inferior parietal lobe.

Modern studies describe deficits in aphasic patients in terms of specific linguistic representations and psycholinguistic processes. This is similar to modern studies using functional neuroimaging. In the next sections, I will review studies of localization of two language processes – spoken word recognition and comprehension, and syntactic operations in comprehension. I have selected these functions because they contrast in several ways – they vary in their abstractness, the extent of the temporal intervals over which they integrate sensory input, and the extent to which current studies converge on localization of the operations they involve. They thus reflect the range of results found in current studies of localization of components of the language processing system.

2.3 Lexical access and word meaning

Many of the processes involved in recognizing linguistically relevant units of sound, activating words from these units, and activating the meanings of words appear to rely on specific areas of the brain; that is, they appear to be narrowly localized. This is not evident in the aphasia literature, however. Nearly all aphasic individuals display speech perception impairments (Caplan, Gow, and Makris 1995; Blumstein 2009; Csepe et al. 2001). The early aphasia literature, examining performance of patients classified according to traditional aphasia syndrome-complexes, showed that Broca's aphasic individuals with lesions involving the inferior frontal gyrus (IFG) and those with Wernicke's aphasia and lesions involving the superior temporal gyrus (STG) displayed deficits in categorically perceiving stimuli along an acoustic–phonetic continuum such as voice-onset-time (VOT) (distinguishing voiced and voiceless stop consonants) and direction of formant transitions (distinguishing place of articulation in stop consonants) (Blumstein, Baker, and Goodglass 1977; Blumstein et al. 1984; Gandour and Dardarananda 1982). Aphasic patients, irrespective of clinical classification, also show impairments in the discrimination of phonological contrasts associated with voicing and place of articulation in stop consonants [p t k b d g] (Basso, Casati, and Vignolo 1977; Blumstein, Baker, and Goodglass 1977; Blumstein et al. 1984; Caplan, Gow, and Makris 1995; Carpenter and Rutherford 1973; Gandour and Dardarananda 1982; Gow and Caplan 1996; Leeper, Shewan, and Booth 1986), although the extent of deficit varies across patients.

However, results from functional neuroimaging studies with normal participants have provided evidence that speech perception recruits a neural processing stream involving both left posterior and anterior brain structures (Scott and Wise 2004; Hickok and Poeppel 2007; Binder and Price 2001) and that this neural stream consists of different processing stages that are located in specific areas. The evidence suggests that temporospectral acoustic cues to feature identity are integrated in the unimodal auditory association cortex lying along the superior temporal sulcus immediately adjacent to the primary auditory koniocortex (Binder et al. 2000). There may also be a difference in the role of the left and right superior temporal sulcus in acoustic–phonetic conversion, with the right hemisphere being more sensitive to longer temporal intervals, and the left to shorter intervals (Hickok and Poeppel 2007). Some researchers have suggested that the unconscious, automatic activation of features and phonemes as a stage in word recognition under normal conditions occurs bilaterally, and that the dominant hemisphere is the sole site only of phonemic processing that is associated with controlled processes such as subvocal rehearsal and conscious processes such as explicit phoneme discrimination and identification, making judgments about rhyme, and other similar functions (Hickok and Poeppel 2004, 2007).

The location of the lexicon is more controversial. Some researchers (e.g., Scott and Johnsrude 2003) have argued that activation of the long term representations of the sound patterns of words occurs in the left superior temporal gyrus. These researchers have argued that there is a pathway along this gyrus and the corresponding left superior temporal sulcus such that word recognition occurs in a region anterior and inferior to primary auditory cortex, and that word meanings are activated further along this pathway in anterior inferior temporal lobe bilaterally (Scott and Johnsrude

2003; Scott and Wise 2004). This pathway constitutes the auditory counterpart to the visual "what" pathway in the inferior occipital-temporal lobe, which recognizes visually presented objects.

The ventral pathway has been distinguished from a "dorsal" pathway that is the counterpart to the system that underlies visual perceptual-motor integrated activity. Hickok and Poeppel (2007) developed the first comprehensive statement of the two-pathway model. They proposed that the ventral stream consists of an acoustic–phonetic conversion process located in STG bilaterally and a "lexico-semantic interface" in the posterior part of the left MTG, which is connected to IFG. The ventral pathway (to MTG) is activated automatically and unconsciously in the process of ordinary spoken word comprehension. The dorsal pathway, which is activated in tasks that involve conscious and/or controlled metaphonological processing, as noted above, transmits phonetic representations from STG to a left dominant temporoparietal sensorimotor interface area (Spt) that interacts reciprocally with an articulatory network comprising the pIFG (posterior inferior frontal gyrus), PM (premotor cortex), and anterior insula. Hickok and Poeppel do not recognize a lexicon in this pathway, but some researchers argue that this pathway also houses lexical representations, in the left supramarginal gyrus (SMG) (Gow et al. 2008; Gow and Segawa 2009) and/or left angular gyrus (AG) (Blumstein 2009). MEG studies have demonstrated complex interactions between auditory–phonetic perceptual processing in STS, lexical activation (in MTG and SMG), and articulatory coding (in PM), that differs as a function of task, with more involvement of the ventral pathway when participants match spoken items to pictures and of the dorsal pathway when they make categorical judgments about auditory stimuli (Gow et al. 2008; Gow and Segawa 2009).

Turning to word meanings, as described above, traditional neurological models maintained that the meanings of words consist of sets of neural correlates of the physical properties that are associated with an auditorily presented word, all converging in the inferior parietal lobe. It is now known that most lesions in the inferior parietal lobe do not affect word meaning and that functional neuroimaging studies designed to require word meaning do not tend to activate this region. Evidence is accruing that the associations of words include "retroactivation" of neural patterns back to unimodal motor and sensory association cortex (Damasio 1989), and that different types of words activate different cortical regions. Verbs are more likely to activate the frontal cortex, and nouns temporal cortex, possibly because verbs refer to actions and nouns refer to static items. More fine-grained relations between words and activation of sensory-motor brain areas have been described (Pulvermüller 2005). Hauk, Johnsrude, and Pulvermüller (2004) have found activation in primary motor and premotor cortex at the homuncular level of the leg when subjects recognized the written word "kick," the mouth when they recognized "lick," and the hand when they recognized "pick."

Both deficits and functional activation studies have suggested that there are unique neural loci for the representation of object categories such as tools (frontal association cortex and middle temporal lobe), animals and foods (inferior temporal lobe and superior temporal sulcus), and faces (fusiform gyrus) (see Caramazza and Mahon 2003, 2006, for review), and some of these localizations are adjacent to areas of the brain that support perception of features that are related to these categories (animals activate an area of lateral temporal adjacent to the area sensitive to biological motion). At the same time as these specializations receive support, evidence from patients with semantic dementia and from functional neuroimaging indicates that another critical part of the semantic network that represents word meanings and concepts is located in the anterior inferior temporal lobes. It has been suggested that this region is a "convergence" area, where different properties of items are bound (Hodges et al. 1992; Devlin, Matthews, and Rushworth 2003).

2.4 Syntactic operations

Deficit–lesion correlations have led to proposals regarding the neural basis for syntactic processing in comprehension. As discussed above, patients are diagnosed as having disorders of

syntactically based comprehension on the basis of retention of the ability to use syntax to understand semantically reversible sentences with simple syntactic structures (which are understood by heuristics) and semantically irreversible sentences with complex syntactic structures (as discussed in Section 1.4), and a disorder of the ability to understand semantically reversible sentences with complex structures. Caramazza and Zurif (1976) noted that some patients with this pattern had "Broca's aphasia," and produced "agrammatic" speech (speech that lacks function words and morphological markers). They suggested that these deficits were connected, and that the underlying deficit in these patients affected syntactic "algorithms," and suggested that these rules were supported by Broca's area (the left pIFG). Variants of this hypothesis have been developed by a number of researchers, in particular Grodzinsky (1990, 1995, 2000), who linked lesions in this area to the loss of quite specific syntactic elements ("traces," in Chomsky's (1986, 1995) model of syntactic structure).

However, the data that are available indicate that deficits in syntactic processing in sentence comprehension occur in all aphasic syndromes and after lesions throughout the perisylvian cortex (Caplan, Baker, and Dehaut 1985; Caplan, Hildebrandt, and Makris 1996; Caramazza et al. 2001). Conversely, patients of all types and with all lesion locations have been described with normal syntactic comprehension (Caplan, Baker, and Dehaut 1985). Caplan, Waters, Kennedy et al. (2007) studied 42 patients with aphasia secondary to left hemisphere strokes and 25 control subjects for the ability to assign and interpret three syntactic structures in enactment, sentence–picture matching, and grammaticality judgment tasks. In regression analyses, lesion measures in both perisylvian and nonperisylvian regions of interest predicted performance after factors such as age, time since stroke, and total lesion volume had been entered into the equations. Patients who performed at similar levels behaviorally had lesions of very different sizes, and patients with equivalent lesion sizes varied greatly in their level of performance. This suggests that localization of these operations is variable.

Functional neuroimaging studies have also led several researchers to suggest that syntactic processing is localized in Broca's area, or in portions of this region (see Grodzinsky and Friederici 2006, for review). However, most neuroimaging studies show that multiple cortical areas are activated in tasks that involve syntactic processing and, as with deficit–lesion correlation studies, different studies have shown different patterns of activation (Caplan 2010).

The discrepancies across studies using both patient data and activation techniques require an explanation. One possibility is that the neural tissue that is responsible for the operations underlying sentence comprehension and syntactic processing is localized in several neural regions, and that the extent to which an area is involved differs in different individuals. This would be consistent with the pattern of deficit–lesion correlations reported by Caplan, Waters, Kennedy et al. (2007) but is hard to reconcile with the neuroimaging results because, if it were true, one would expect there to be a variety of activation patterns in the individuals tested in any one study, most likely resulting in no specific activation in most studies. Another possibility is that different studies have involved different aspects of syntactic processing. As illustrated in the discussion of online performance, poor performance in an aphasic patient can reflect many different deficits. Different patterns of deficit–lesion correlation in different studies are consistent with localization of different affected processes in different patients in different brain regions. The variety of activation patterns seen in functional neuroimaging may have a similar explanation. Different sentence contrasts in different studies highlight different syntactic operations and may activate correspondingly different brain areas. In addition, no one experimental contrast isolates a single operation, so some activation in most studies is likely due to activation of associated operations. Similar to the finding of task- and sentence-type-specific deficits in aphasic patients, the same sentence contrasts have been shown to activate different areas in different tasks (Caplan 2010), indicating that some activation reflects the interaction of comprehension with satisfying the demands of different tasks.

On the other hand, different types of linguistic operations may be supported by different types of brain organization. For instance, it is possible that perceptual identification of specific

items is invariably carried out by small cortical areas closely linked to primary perceptual areas, and that more abstract operations, such as those related to syntactic analysis, are carried out by multiple areas that do show some variability across the population. These are questions for further research.

3 Conclusion

Contemporary studies of language disorders and the relationship of language to the brain are beginning to deal with difficult questions such as: What is the nature of disorders of linguistic representations and their processing? Where are these representations and processes localized in the brain? Research into these questions is now at a point where we can propose hypotheses and support (and attack) them empirically. This is an important advance over the level of description and modeling of language disorders and the neural basis for language that characterized the field for decades. I have outlined a small set of studies that characterize this field. The reader will appreciate that these are new questions, being approached with new psycholinguistic paradigms and new technologies to describe the brain, and there are no sure answers at present to these basic questions. Nonetheless, research is clearly in the direction of a more detailed and accurate understanding of language disorders and the neural mechanisms that support normal language knowledge and use.

RELEVANT JOURNALS AND SOCIETIES

Journals
Aphasiology
Brain and Language
Journal of Speech, Language, Hearing Disorders
Journal of Neurolinguistics Academy of Aphasia

Societies
Academy of Aphasia
Neurobiology of Language
Science of Aphasia

EMERGING TRENDS AND RESEARCH QUESTIONS

1. Online studies of aphasic deficits
2. Neuroimaging of language processing in aphasia
3. Developing techniques for neuroimaging of language processing (MEG, DTI, DSI, receptor mapping)
4. Genetic expression and language
5. TMS and other stimulation techniques applied to language

REFERENCES

Annett, M. (1985). *Left, Right, Hand and Brain: The Right Shift Theory*. London: Erlbaum.

Basso, A., Casati, G., and Vignolo, L. A. (1977). Phonemic identification defects in aphasia. *Cortex*, 13, 84–95.

Basso, A., Lecours Roch, A., Moraschini, S., and Vanier, M. (1985). Anatomoclinical correlations of the aphasias as defined through computerized tomography: exceptions. *Brain and Language*, 26, 201–29.

Beeman, M. (1998). Coarse semantic coding and discourse comprehension. In M. Beeman and C. Chiarello (eds), *Right Hemisphere Language Comprehension: Perspectives from Cognitive Neuroscience* (pp. 255–84). Mahwah, NJ: Erlbaum.

Ben-Shachar, M., Palti, D., and Grodzinsky, Y. (2004). The neural correlates of syntactic movement: converging evidence from two fMRI experiments. *Neuroimage*, 21, 1320–36.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., … Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10, 512–28.

Binder, J. R. and Price, C. J. (2001). Functional neuroimaging of language. In R. Cabeza and A. Kingstone (eds), *Handbook of Functional Neuroimaging.* Cambridge, MA: MIT Press.

Blumstein, S. E. (2009). Auditory word recognition: evidence from aphasia and functional neuroimaging. *Language and Linguistic Compass*, 3, 824–38.

Blumstein S. E., Baker, E., and Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, 15, 19–30.

Blumstein, S., Byma, G. et al. (1998). On-line processing of filler-gap constructions in aphasia. *Brain and Language*, 61, 2, 149–69.

Blumstein, S. E., Milberg, W., and Shrier, R. (1982). Semantic processing in aphasia: evidence from an auditory lexical decision task. *Brain and Language*, 17, 301–15.

Blumstein S. E., Tartter V. C., Nigro, G., and Statlender, S. (1984). Acoustic cues for the perception of place of articulation in aphasia. *Brain and Language*, 22, 128–49.

Broca, P. (1861). Remarques sur le siège de la faculté de la parole articulée, suives d'une observation d'aphémie (perte de parole). *Bulletin de la Société d'Anatomie*, 36, 330–57.

Brown, J. (1972). *Aphasia, Apraxia and Agnosia – Clinical and Theoretical Aspects*. Springfield, IL: Thomas.

Brownell, H. H. and Gardner, H. (1988). Neuropsychological insights into humour. In J. Durant and J. Miller (eds), *Laughing Matters* (pp. 17–34). Harlow: Longman Scientific.

Bub, D. N., Black, S., Hampson, E., and Kertesz, A. (1988). Semantic encoding of pictures and words: some neuropsychological observations. *Cognitive Neuropsychology*, 5, 1, 27–66.

Butterworth, B., Campbell, R., and Howard, D. (1986). The uses of short-term memory: a case study. *Quarterly Journal of Experimental Psychology*, 38, 705–37.

Caplan, D. (1992). *Language: Structure, Processing, and Disorders*. Cambridge, MA: MIT Press.

Caplan, D. (2010). Task effects on BOLD signal correlates of implicit syntactic processing. *Language and Cognitive Processes*, 25, 866–901.

Caplan, D., Baker, C., and Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition, 21*, 117–75.

Caplan, D., DeDe. G., and Michaud, J. (2006) Task-independent and task-specific syntactic deficits in aphasic comprehension, *Aphasiology*, 20, 893–920.

Caplan, D. and Futter, C. (1986). Assignment of thematic roles to nouns in sentence comprehension by an agrammatic patient. *Brain and Language*, 27, 117–34.

Caplan, D., Gow, D., and Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic–phonetic processing deficits. *Neurology*, 45, 293–8.

Caplan, D. and Hildebrandt, H. (1988). *Disorders of Syntactic Comprehension*. Cambridge, MA: MIT Press.

Caplan, D., Hildebrandt, N., and Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain*, 119, 933–49.

Caplan, D. and Waters, G. (1990). Short-term memory and language comprehension: a critical review of the neuropsychological literature In T. Shallice and G. Vallar (eds), *The Neuropsychology of Short-Term Memory* (pp. 337–89). Cambridge: Cambridge University Press.

Caplan, D. and Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22, 114–26.

Caplan, D. and Waters, G. S. (2003). On-line syntactic processing in aphasia: studies with auditory moving windows presentation. *Brain and Language*, 84, 2, 222–49.

Caplan, D., Waters, G. S., DeDe, G., Michaud, J., and Reddy, A. (2007). A study of syntactic processing in aphasia I: behavioral (psycholinguistic) aspects. *Brain and Language*, 101, 2, 103–50.

Caplan, D., Waters, G. S., and Hildebrandt, N. (1997): Syntactic determinants of sentence comprehension in aphasic patients in sentence–picture matching tests. *Journal of Speech and Hearing Research*, 40, 542–55.

Caplan, D., Waters, G., Kennedy, D., Alpert, N. Makris, N., … Reddy, A. (2007). A study of syntactic processing in aphasia II: neurological aspects. *Brain and Language*, 101, 2, 150–71.

Cappa, S., Cavalotti, G., Guidotti, N., Papagno, C., and Vignolo, L. (1983). Subcortical aphasia: two clinical CT scan correlation studies. *Cortex*, 19, 227–41.

Caramazza, A., Capitani, E., Rey, A., and Berndt, R. S. (2001). Agrammatic Broca's aphasia is not associated with a single pattern of comprehension performance. *Brain and Language*, 76, 158–84.

Caramazza, A., Hillis, A. E., Rapp, B. C., and Romani, C. (1990). The multiple semantics hypothesis:

multiple confusions? *Cognitive Neuropsychology*, 7, 3, 161–89.

Caramazza, A. and Mahon, B. Z. (2003). The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, 7, 325–74.

Caramazza, A., and Mahon, B. Z. (2006). The organization of conceptual knowledge in the brain: the future's past and some future directions. *Cognitive Neuropsychology*, 23, 1, 13–38.

Caramazza, A. and Shelton, J. R. (1998). Domain specific knowledge systems in the brain: the animate–inanimate distinction. *Journal of Cognitive Neuroscience*, 10, 1–34.

Caramazza, A. and Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, 3, 572–82.

Carpenter, R. L. and Rutherford, D. R. (1973). Acoustic cue discrimination in adult aphasia. *Journal of Speech and Hearing Research*, 16, 534–44.

Chiarello, C., Burgess, C., Richards, L., and Pollock, A. (1990). Semantic and associative priming in the cerebral hemispheres: some words do, some words don't...sometimes, some places. *Brain and Language, 38*, 75–104.

Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.

Chomsky, N. (1995). *Barriers*. Cambridge, MA: MIT Press.

Cloutman, L., Newhart, M., Davis, C., Heidler-Gary, J., and Hillis, A. E. (2009). Acute recovery of oral word production following stroke: patterns of performance as predictors of recovery. *Behavioral Neurology*, 21, 3, 145–53.

Cree, G. S. and McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, 132, 2, 163–201.

Crosson, B. (1985). Subcortical functions in language: a working model. *Brain and Language*, 25, 257–92.

Crutch, S. J. and Warrington, E. K. (2003). The selective impairment of fruit and vegetable knowledge: a multiple processing channels account of fine-grain category specificity. *Cognitive Neuropsychology*, 20, 355–73.

Csepe, V., Osman-Sagi, J., Molnar, M., and Gosy, M. (2001). Impaired speech perception in aphasic patients: event-related potential and neuropsychological assessment. *Neuropsychologia*, 39, 1194–208.

Cupples, L. and Inglis, A. L. (1993) When task demands induce "asyntactic" comprehension: A study of sentence interpretation in aphasia, *Cognitive Neuropsychology*, 10, 201– 234

Damasio, A. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33, 25–62.

Damasio, A., Damasio, H., Rizzo, M., Varney, N., and Gersch, F. (1982). Aphasia with nonhemmorhagic lesions in the basal ganglia and internal capsule. *Archives of Neurology*, 39, 15–20.

Dehaene, S. and Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Science*, 15, 254–62.

Devlin, J. T., Gonnerman, L. M., Anderson, E. S., and Seidenberg, M. S. (1998). Category-specific semantic deficits in focal and widespread brain damage: a computational account. *Journal of Cognitive Neuroscience*, 10, 77–94.

Devlin, J. T., Matthews, P. M., and Rushworth, M. F. S. (2003). Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience*, 15, 71–84.

Dickey, M. W. and Thompson, C. K. (2004). The resolution and recovery of filler-gap dependencies in aphasia: evidence from on-line anomaly detection. *Brain and Language*, 88, 1, 108–27.

Friederici, A. and Kilborn, K. (1989). Temporal constraints on language processing: syntactic priming in Broca's aphasia. *Journal of Cognitive Neuroscience*, 1, 262–72.

Gandour J. and Dardarananda R. (1982) Voice onset time in aphasia: Thai, I. Perception. *Brain and Language*, 17, 24–33.

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., and Patterson, K. (2001). Prototypicality, distinctiveness and intercorrelation: analyses of semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology*, 18, 125–74.

Gazzaniga, M. S. (1983). Right hemisphere language following brain bisection: a 20-year perspective. *American Psychologist*, 38, 525–49.

Geschwind, N. (1965). Disconnection syndromes in animals and man. *Brain*, 88, 237–94, 585–644.

Geschwind, N. and Galaburda, A. (1985). Cerebral lateralization: biological mechanism, associations and pathology i–iii: a hypothesis and a program for research. *Archives of Neurology*, 42, 428–59(i), 421–52(ii), 634–54(iii).

Geschwind, N. and Galaburda, A. (1987). *Cerebral Lateralization: Biological Mechanisms, Associations and Pathology*. Cambridge, MA: MIT Press.

Gonnerman, L., Andersen, E. S., Devlin, J. T., Kempler, D., and Seidenberg, M. (1997). Double

dis- sociation of semantic categories in Alzheimer's disease. *Brain and Language*, 57, 254–79.

Goodglass, H. and Quadfasel, F. A. (1954). Language laterality in left-handed aphasics. *Brain*, 77, 521–48.

Gow, D. W., Jr. and Caplan, D. (1996) An examination of impaired acoustic–phonetic processing in aphasia. *Brain and Language*, 52, 386–407.

Gow, D. W. and Segawa, J. A. (2009). Articulatory mediation of speech perception: a causal analysis of multi-modal imaging data. *Cognition*, 110, 222–36.

Gow, D. W., Segawa, J. A., Alfhors, S., and Lin, F-H. (2008). Lexical influences on speech perception: a Granger causality analysis of MEG and EEG source estimates. *NeuroImage*, 43, 614–23.

Grodzinsky, Y. (1990). *Theoretical Perspectives on Language Deficits*. Cambridge, MA: MIT Press.

Grodzinsky, Y. (1995). A restrictive theory of agrammatic comprehension. *Brain and Language*, 50, 27–51.

Grodzinsky, Y. (2000). The neurology of syntax: language use without Broca's area. *Behavioral and Brain Sciences*, 23, 47–117.

Grodzinsky, Y., and Friederici, A. D. (2006). Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology*, *16*(2), 240–246

Haarmann, H. J. and Kolk, H. H. J. (1991). Syntactic priming in Broca's aphasics: evidence for slow syntax. *Aphasiology*, 5, 247–63.

Hauk, O., Johnsrude, I., and Pulvermüller, F (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41, 301–7.

Hickok, G. and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67–99.

Hickok, G. and Poeppel, D. (2007) The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402.

Hildebrandt, N., Caplan, D., and Evans, K. (1987). The mani lefti without a trace: a case study of aphasic processing of empty categories. *Cognitive Neuropsychology*, 4, 3, 257–302.

Hillis, A., Rapp, B., Romani, C., and Caramazza, A. (1990). Selective impairment of semantics in lexical processing. *Cognitive Neuropsychology*, 7, 3, 191–243.

Hodges, J., Patterson, K., Oxbury, S., and Funnell, E. (1992). Semantic dementia. Progressive fluent aphasia with temporal lobe atrophy. *Brain*, 115, 1783–806.

Howard, D. and Orchard-Lisle, V. (1984). On the origin of semantic errors in naming: evidence from the case of a global aphasic. *Cognitive Neuropsychology*, 1, 163–90.

Humphreys, G. W. and Forde, E. M. (2001). Hierarchies, similarity, and interactivity in object recognition: "category-specific" neuropsychological deficits. *Behavioral and Brain Sciences*, 24, 453–509.

Joanette, Y. and Brownell, H. H. (1990). *Discourse Ability and Brain Damage: Theoretical and Empirical Perspectives*. New York: Springer.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., and Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, 274, 114–16.

Kempler, D., Metter, E., Jackson, C., Hanson, W., Riege, W., … Phelps, M. (1988). Disconnection and cerebral metabolism: the case of conduction aphasia. *Archives of Neurology*, 45, 275–9.

Kinsbourne, M. (1971). The minor cerebral hemisphere as a source of aphasic speech. *Archives of Neurology*, 25, 302–6.

Kolb, M. and Whishaw, I. Q. (2003). *Fundamentals of Human Neuropsychology* (7th edn). New York: Worth.

Leeper, H. A., Jr., Shewan, C., and Booth, J. C. (1986) Altered acoustic cue discrimination in Broca's and Conduction aphasics. *Journal of Communication Disorders*, 19, 83–103.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Lichtheim, L. (1885). Brain: on aphasia. *Abstracts of British and Foreign Journals*, 433–84.

Linebarger, M. C. (1990). Neuropsychology of sentence parsing. In A. Caramazza (ed.), *Cognitive Neuropsychology and Neurolinguistics: Advances in Models of Cognitive Function and Impairment* (pp. 55–122). Hillsdale, NJ: Erlbaum.

Linebarger, M. C. (1995). Agrammatism as evidence about grammar. *Brain and Language*, 50, 52–91.

Linebarger, M. C., Schwartz, M. F., and Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, 13, 361–92.

Love, T., Swinney, D. and Zurif, E. (2001). Aphasia and the time-course of processing long distance dependencies. *Brain and Language*, 79, 169–70.

Love, T., Swinney, D., Walenski, M., and Zurif, E. (2008) How left inferior frontal cortex participates in syntactic processing: evidence from aphasia. *Brain and Language*, 107, 203–19.

Luria, A. R. (1970) *Traumatic Aphasia*. The Hague: Mouton de Gruyter.

Martin, R. C. and He, T. (2004). Semantic short-term memory and its role in sentence processing: a replication. *Brain and Language*, 89, 76–82.

Masdeu, J. C., Schoene, W. C., and Funkenstein, H. (1978). Aphasia following infarction of the left supplementary motor area: a clinicopathological study. *Neurology*, 28, 220–3.

McCarthy, R. A. and Warrington, E. K. (1984). A two-route model of speech production: evidence from aphasia. *Brain*, 107, 463–85.

McGlone, J. (1980). Sex differences in human brain asymmetry: a critical survey. *Behavioral and Brain Sciences*, 3, 215–63.

McRae, K. and Cree, G. S. (2002). Factors underlying category-specific semantic impairments. In E. M. E. Forde and G. W. Humphreys (eds), *Category-Specificity in the Brain and Mind* (pp. 211–48). New York: Psychology Press.

Metter, E. J., Kempler, D., Jackson, C. A., Hanson, W. R., Riege, W. H., … Phelps, M. E. (1987). Cerebellar glucose metabolism and chronic aphasia. *Neurology*, 37, 1599–606.

Metter, E., Riege, W., Hanson, W., Jackson, C., Kempler, D., and VanLancker, D. (1983). Comparison of metabolic rates, language and memory, and subcortical aphasias. *Brain and Language*, 19, 33–47.

Metter, E., Riege, W., Hanson, W., Jackson, C., Kempler, D., and VanLancker, D. (1988). Subcortical structures in aphasia: an analysis based on (F-18)-fluorodeoxyglucose positron emission tomography and computed tomography. *Archives of Neurology*, 45, 1229–34.

Miceli, G., Silveri, M., Villa, G., and Caramazza, A. (1984). On the basis for the agrammatic's difficulty in producing main verbs. *Cortex*, 20, 207–20.

Milberg, W. and Blumstein, S. E. (1981). Lexical decision and aphasia: evidence for semantic processing. *Brain and Language*, 14, 371–85.

Miller, G. A. and Johnson-Laird, P. N. (1976). *Language and Perception*. Cambridge, MA: Harvard University Press.

Milner, B. (1974). Hemispheric specialization: its scope and limits. In F. O. Schmitt and F. G. Worden (eds), *The Neurosciences: Third Study Program* (pp. 75–89). Cambridge, MA: MIT Press.

Milner, B., Branch, C., and Rasmussen, T. (1964). Observations on cerebral dominance. In A. de Reuck and M. O'Connor (eds), *Disorders of Language* (pp. 200–14). London: J. & A. Churchill.

Mohr, J., Watters, W., and Duncan, G. (1975). Thalamic hemorrage and aphasia. *Brain and Language*, 2, 3–17.

Molloy, R., Brownell, H. H., and Gardner, H. (1990). Discourse comprehension by right-hemisphere stroke patients: deficits of prediction and revision. In Y. Joanette and H. H. Brownell (eds), *Discourse Ability and Brain Damage: Theoretical and Empirical Perspectives* (pp. 113–30). New York: Springer.

Naeser, M. A., Alexander, M. P., Helm-Estabrooks, N., Levine, H. L., Laughlin, S. A., and Geschwind, N. (1982). Aphasia with predominantly subcortical lesion sites: description of three capsular/putaminal aphasia syndromes. *Archives of Neurology*, 39, 2–14.

Olsen, T. S., Bruhn, P., and Oberg, R. (1986). Cortical hypertension as a possible cause of subcortical aphasia. *Brain*, 109, 393–410.

Paivio, A. (1971). *Imagery and Verbal Processes*. London: Holt, Rinehart & Winston.

Perani, D., Vallar, G., Cappa, S., Messa, C., and Fazio, F. (1987). Aphasia and neglect after cortical stroke. A clinical/cerebral study. *Brain*, 110, 1211–29.

Pick, A. (1973). *Aphasia*. Springfield, IL: Thomas.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6, 7, 576–82.

Riddoch, M. J. and Humphreys, G. W. (1987). Visual object processing in optic aphasia: a case of semantic access agnosia. *Cognitive Neuropsychology*, 4, 2, 131–85.

Roman, M., Brownell, H. H., Potter, H. H., Seibold, M. S., and Gardner, H. (1987). Script knowledge in right hemisphere-damaged and in normal elderly adults. *Brain and Language*, 31, 151–70.

Russell, W. R. and Esper, M. L. E. (1961). *Traumatic Aphasia*. London: Oxford University Press.

Sartori, G. and Job, R. (1988). The oyster with four legs: a neuropsychological study on the interaction of visual and semantic information. *Cognitive Neuropsychology*, 5, 105–32.

Schwartz, M. (1984). What the classical aphasia categories can't do for us, and why. *Brain and Language*, 21, 1–8.

Schwartz, M., Saffran, E., and Marin, O. (1980). The word order problem in agrammatism I: comprehension. *Brain and Language*, 10, 249–62.

Schwartz, M. F., Linebarger, M. C., and Saffran, E. M. (1985). The status of the syntactic deficit theory of agrammatism. In M.-L. Kean (ed.), *Agrammatism* (pp. 83–124). New York: Academic Press.

Schwartz, M. F., Marin, O. S. M., and Saffran, E. M. (1979). Dissociations of language function in dementia: a case study. *Brain and Language*, 7, 277–306.

Scott, S. K. and Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neuroscience*, 26, 100–7.

Scott, S. K. and Wise, R. J. S. (2004) The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, 92, 13–45.

Shallice, T. (1988a). *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.

Shallice, T. (1988b). Specialisation within the semantic system. *Cognitive Neuropsychology*, 5, 133–42.

Shelton, J. R., Fouch, E., and Caramazza, A. (1998). The selective sparing of body part knowledge: a case study. *Neurocase*, 4, 339–51.

Silveri, M. C. and Gainotti, G. B. (1988). Interaction between vision and language in category specific

semantic impairment for living things. *Cognitive Neuropsychology*, 5, 677–709.

Smith, E. E. and Medin, D. L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Snodgrass, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning and Verbal Behavior*, 23, 3–22.

Subirana, A. (1964). The relationship between handedness and language function. *International Journal of Neurology*, 4, 215–34.

Swinney, D. and Zurif, E. (1995). Syntactic processing in aphasia. *Brain and Language*, 50, 225–39.

Swinney, D., Zurif, E., and Nicol, J. (1989). The effects of focal brain damage on sentence processing: an examination of the neurological organization of a mental module. *Journal of Cognitive Neuroscience*, 1, 25–37.

Thompson, C. K. and Choy, J. (2009). Pronominal resolution and gap filling in agrammatic aphasia: evidence from eye movements. *Journal of Psycholinguistic Research*, 38, 3, 255–83.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving and W. Donaldson (eds), *Organization of Memory* (pp. 381–403). New York: Academic Press.

Tulving, E. (1983). *Elements of Episodic Memory*. Oxford: Oxford University Press.

Tyler, L. (1985). Real-time comprehension processes in agrammatism: a case study. *Brain and Language*, 26, 259–75.

Tyler, L. K. and Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Science*, 5, 244–52.

Tyler, L. K., Moss, H. E., Durrant-Peatfi eld, M. R., and Levy, J. P. (2000). Conceptual structure and the structure of concepts: a distributed account of category-specific deficits. *Brain and Language*, 75, 195–231.

Vinson, D. P., Vigliocco, V., Cappa, S., and Siri, S. (2003). The breakdown of semantic knowledge alone: insights from a statistical model of meaning representation. *Brain and Language*, 86, 3, 347–441.

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, 27, 635–57.

Warrington, E. K. (1981a). Concrete word dyslexia. *British Journal of Psychology*, 72, 175–96.

Warrington, E. K. (1981b). Neuropsychological studies of verbal semantic systems. *Phil. Trans. R. Soc. Lond., B295*, 411–23.

Warrington, E. K. and McCarthy, R. (1983). Category specific access dysphasia. *Brain*, 106, 859–78.

Warrington, E. K. and McCarthy, R. (1987). Categories of knowledge: further fractionation and an attempted integration. *Brain*, 110, 1273–96.

Warrington, E. K. and Shallice, T. (1979). Semantic access dyslexia. *Brain*, 102, 43–63.

Warrington, E. K. and Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107, 829–53.

Waters, G. S., Caplan, D., and Hildebrandt, N. (1991). On the structure of verbal short-term memory and its functional role in sentence comprehension: evidence from neuropsychology. *Cognitive Neuropsychology*, 8, 2, 81–126.

Weisenberg, T. and McBride, K. (1935). *Aphasia*. New York: Commonwealth Fund.

Zingeser, L. B. and Berndt, R. S. (1988). Grammatical class and context effects in a case of pure anomia: implications for models of lexical processing. *Cognitive Neuropsychology*, 4, 473–516.

Zurif, E., Swinney, D., Prather, P., Solomon, J., and Bushell, C. (1993). An on-line analysis of syntactic processing in Broca's and Wernicke's aphasia. *Brain and Language*, 45, 448–64.

17 Psycholinguistics

KIEL CHRISTIANSON

"Neither can embellishments of language be found without arrangement and expression of thoughts, nor can thoughts be made to shine without the light of language."

—Cicero

Human language is seated in each and every normally developed, neurologically healthy human mind. It is the outward expression of language, however, with which most people are familiar: language comprehension and production during the course of spoken, signed, or written communication. Noam Chomsky has often said that if people could simply beam thought directly into each other's brains, we would have no need for language, nor for the intricacies of phonology, morphology, syntax, and the other subcomponents of language elucidated in the chapters of this volume. As it happens, however, language use requires the formal structure of language to interface with other cognitive functions, and these interfaces operate under certain constraints. For example, although language and communication can be dissociated, language is generally used for communication. Because we do not in fact have the aforementioned ability to communicate via telepathic thought, language is temporally constrained. The mouths of speakers and hands of signers are limited in their mobility, and are therefore able to produce essentially one sound, word, or sign at a time. Language production is therefore inherently incremental and serial. Consequently, language comprehension is also incremental, more or less yoked to the unfolding input, be it aural or visual. Psycholinguistics is the study of the interfaces between language structure and language use: between thought and production, production and comprehension, and comprehension and thought, as well as the many other cognitive systems that shape the use of language as it is flowing into and out of the mind in real time. I begin with a (very) brief, quite recent history of the field. In actuality, psychologists have been interested in language since the very inception of psychology as a discipline, as detailed in fascinating detail by Levelt (2012). The "Cognitive Revolution" in the 1950s (Miller 2003) led to closer ties between the psychological study of language and the linguistic study of language, and this is where the following brief chapter will begin. We will then focus on a few central issues in comprehension and production, concluding with a (small) sampling of psycholinguistic research examining how these capacities develop both in children as they learn their native languages and in adults learning foreign languages.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

1 Psycholinguistics as a Field of Study

Fernanda Ferreira aptly describes the relationship between formal linguistics and psycholinguistics as a complicated "family relationship" (2005: 365), and interested readers are encouraged to refer to Ferreira's accessible outline of how the two fields have complemented each other over the past 60 plus years since the Cognitive Revolution (and her advice as to how to make "family reunions" somewhat less awkward). Space limitations preclude the inclusion of a satisfactory account of the historical record here. Common linguistic and cognitive threads are woven throughout the history of psycholinguistics, however, which are worth noting. One is the tension between linguistic theories on one hand, and the capacity-limited processing characteristics of the human mind. While the former seek to capture crosslinguistic generalities in the most perspicuous and elegant way possible, the latter must necessarily view these theories through the lens of actual human performance, and, as noted above, performance entails certain constraints. For example, whereas much generative syntactic theory conceives of syntactic structure as being derived essentially from the "middle out" (e.g., Government and Binding Theory: Chomsky 1981; Minimalism: Chomsky 1995; Epstein and Seely 2006), language unfolds over time "left to right."

This theoretical difference in perspective is reflected in the tension in psycholinguistics between "bottom-up" processing and "top-down" processing (e.g., Gibson 2006; see Clark 2013, for an excellent discussion of this tension in all aspects of cognition). Bottom-up describes processing that begins with the lowest level representations available, say, phonemes, and takes the output of that processing level as the input of the next level, say, syllables or morphemes. In strictly serial, bottom-up processing, higher levels of representation should have no influence on lower levels. Top-down describes processing that works in the opposite direction: higher levels of information are used to predict the content of upcoming lower level input and perhaps influence its ongoing integration. The bottom-up aspect of language processing may seem more intuitive at first blush; for example, given a picture of a boy, several toys, and a cake, people are not terribly accurate at predicting the upcoming word in the sentence "*The boy will move the* ___," as measured by the timing of looks in eye movement monitoring experiments to the cake compared to timing of looks to the other objects in the picture (Altmann and Kamide 1999). Yet top-down effects, upon reflection, are also readily apparent. Reicher (1969) documented a "word-superiority effect" for letter identification, such that it is easier to recognize letters when they are presented in words compared to when they are presented in nonword strings or by themselves. Similarly, imagine a slight revision of the sentence above: "*The boy will eat the* ___." People are more likely to look at the cake sooner with this more specific context, in which the cake is the only object in the picture that is edible (Altmann and Kamide 1999; cf. Tanenhaus et al. 1995). Reading times on more predictable words have also been shown to be shorter than on less predictable words, controlling for word length and frequency (Rayner and Well 1996), suggesting that less computational effort is required to identify predictable words during reading (Rayner 1998). Furthermore, predictions about upcoming words during reading do not seem to contain only rough semantic information. Instead, these predictions include, when the context is narrow enough, information about phonological characteristics of the word (Laszlo and Federmeier 2009) and even letter order (Luke and Christianson 2012).

The present era of psycholinguistic research is particularly exciting, as this tension, which has been present from the days when psycholinguistics emerged as a unique field of study, is being addressed with the aid of powerful new methods, including brain imaging and computational modeling. The remainder of this chapter will highlight some of these efforts, and point toward future research that would appear to be logical extensions of today's most thought-provoking work.

2 Language Production

Thought is instantaneous and holistic, but speech is necessarily sequential and segmental. A great deal of research in language production has sought to clarify how amorphous thought gets squeezed out through the bottleneck of linear speech. The most influential model of speech production was developed by Levelt (1998; Levelt, Roelofs, and Meyer 1999), and it has generated a great deal of research ever since. Within this model, a message is first generated and associated with lexicalized concepts. Next, the message passes on to a "formulator," whose job it is to determine the grammatical properties of the lexicalized concepts and associate those concepts with lemmas (word forms) from the lexicon, including appropriate morphological forms. Next, the phonological properties of the lemmas are filled in, and finally, articulation commences.

It is clear from this brief sketch of the Levelt production model that there is considerable leeway as to the amount of planning that may occur in speech production. Returning to the tension between bottom-up and top-down processing, one might imagine a model of speech production in which speakers simply produce the first, most conceptually accessible or salient entity that comes to mind and then try to fill in a licit syntactic structure from there. Such a system has been termed "radically incremental" (Ferreira and Swets 2002). Although such a model has been proposed to account for some production data (e.g., De Smedt 1990; Kempen and Hoenkamp 1987; van Nice and Dietrich 2003), it has generally been dismissed (e.g., Bock 1986) as unrealistic. Difficulty arises, however, in decisively ruling out strictly word-by-word incrementality in languages like English, in which conceptual accessibility, syntactic structure, and linear order are badly confounded. For example, Bock and Warren (1985) demonstrated that if people read the word *worship* and are then shown a picture of lightning striking a church, they are more likely to describe the church in the passive voice (*The church is struck by lightning*) than if they had previously read a word that was semantically unrelated to the concept "church." Bock and Warren interpreted this finding as showing that the most conceptually accessible entity claims not only the first position in the string, but also the most prominent syntactic position. As noted, though, English is not a particularly good language in which to examine these issues, as subjects tend to come first in English sentences, and both discourse topics (i.e., "given" information) and thematic agents (i.e., "doers" of actions) tend to be subjects. Furthermore, English does not have any special morphological marking for discourse topics or thematic agents, or syntactic subjects, for that matter.

As it turns out, other Indo-European languages are not much better suited as testing grounds for language production models (Levelt 1998). Recently, researchers have begun to examine production in non-Indo-European languages with the aim of disentangling syntactic, semantic, and discourse factors affecting language production (e.g., Christianson and F. Ferreira 2005; V. Ferreira and Yoshita 2003; Yamashita and Chang 2001). The study by Christianson and F. Ferreira allowed for the dissociation of these factors by examining production in the indigenous North American Algonquian language Odawa. Without going into detail about the morphosyntax of Odawa, suffice it to say that the language has a number of characteristics (Valentine 2001) that make it intriguingly dissimilar from languages previously studied by psycholinguists: a broad inventory of verb forms, a complex agreement system of nominal and verbal morphology that is not technically case marking, pro-drop (the omission of noun arguments), and extremely flexible word order (all six logical word orders – SOV, OSV, SVO, OVS, VSO, VOS – are possible and are used). These features combine in Odawa to make it possible to utter the word for any conceptually accessible entity first without having to change the underlying syntax or to use a less frequent verb form, both of which are the case in English. The results from Odawa, however, point to mechanisms of language production that are similar to English, despite the typological distance between the two languages. Odawa speakers altered the syntactic structures they used to describe pictures in response to question prompts that highlighted either the thematic agent,

thematic patient, or neither; they did not just spit out the first word that came to mind. Taken together, the Odawa results supported conclusions drawn by V. Ferreira and Dell (2000), who proposed that production is not a figurative "race" out of the mouth. Instead, syntactic structures are chosen to accommodate highly activated concepts at the message level; if a highly activated concept is a thematic patient, that property mediates the selection in English of a passive structure, which serves to align the thematic, syntactic, and discourse status of the entity with the first position in the utterance. This view of language production can be termed "weakly incremental."

Related evidence of syntactic effects on language production can be found in the extensive literature on syntactic, or structural, priming. This literature is based on the observation by Bock (1986) that people tend to repeat recently heard or spoken syntactic structures to encode new information. For example, if you happen to hear a sentence such as *The president gave the Congress an ultimatum*, you are more likely to say *The girl handed the boy a flower* than the meaning-equivalent alternative *The girl handed a flower to the boy* to describe a picture of a girl handing a flower to a boy. The literature exploring this priming effect is enormous, and well beyond the scope of this chapter to even summarize. The interested reader is encouraged to refer to an excellent review of structural priming by Pickering and V. Ferreira (2008); however, a few intriguing findings are worth noting here. It is possible that explicit memory for the content or structure of a sentence results in structural priming (Bock and Griffin 2000; Chang, Dell, and Bock 2006; Hartsuiker et al. 2008), as suggested by the observation that priming rates are "boosted" when verbs are repeated from the prime in the production (e.g., Hartsuiker et al. 2008). The possibility that this is the only mechanism underlying structural priming appears to be ruled out, however, by evidence that priming persists over up to ten intervening sentences (Bock and Griffin 2000). More strikingly, structural priming effects have been observed in amnesic patients who have no explicit memory for the prime sentences (V. Ferreira et al. 2008). Finally, evidence is accruing that abstract syntactic structure is shared between languages in the minds of bilinguals despite the lack of lexical or even word order overlap between the two languages (e.g., Korean and English) (Shin and Christianson 2009, 2011; cf. Chen et al. 2013; Hartsuiker, Pickering, and Veltkamp 2004; Kantola and van Gompel 2011).

Language production thus appears to involve a degree of planning (e.g., Garrett 1975; Mackay 1972), but it is important to note that no one believes entire sentences are planned out before speaking commences in normal conversational settings. Although it is not yet known for certain what the range of planning is (see Konopka 2012), a good approximation appears to be more or less one clause. Speech error data provide evidence along with the studies cited above that converge on the primacy of the clause in language production planning. Consider this speech error, reported in Jaeger (2007): *My sore is really … My throat is really sore*. This is an example of an anticipatory error in that a word farther along in the message is produced too soon. It is typical of anticipatory speech errors, the vast majority of which occur inside a single clause (Garrett 1975). On the other hand, it is a bit unusual in that an adjective (*sore*) is produced in place of a noun (*throat*). Perhaps the error signal associated with this cross-category swap alerted the speaker that something was wrong so that the utterance was interrupted and corrected mid-sentence. An example of a complete exchange between two nouns, reported by Garrett (1975), was not interrupted in this way: *Stop beating your brick against a head wall*. Errors in which substituted words and intended words are from the same category are far more common than errors in which words from different categories are exchanged (Dell 1986; Postma 2000).

Before closing this section, it is important to note the growing body of literature on language production in more naturalistic tasks, such as actual conversations between two people. This research has demonstrated that the linguistic expression of a message is affected significantly by elements of preceding discourse, the visual information in the immediate environment, and the perceived joint attention of both speaker and listener. Speakers in a conversation tend to align referential forms, using the same, often tacitly agreed upon, terms for entities (Brennan and Clark 1996; Clark and Wilkes-Gibbs 1989; Wilkes-Gibbs and Clark 1992). They modify referential

expressions according to task demands and likelihood of confusion on the part of the listener (e.g., Brown-Schmidt, Campana, and Tanenhaus 2005; Brown-Schmidt and Tanenhaus 2008). Some research has also shown that speakers monitor their ongoing beliefs about the ease of comprehension on the part of addressees and modify the syntactic structure of utterances in order to avoid ambiguities (Haywood, Pickering, and Branigan 2005). In short, speakers in conversational settings appear to adjust their utterances on the fly as dictated by their perception of the listener's comprehension and the contents of the environment – a strikingly top-down influence on language production.

Speakers are not perfect at doing this, however. In contrast to Haywood, Pickering, and Branigan's (2005) results, neither Arnold et al. (2004) nor Kraljic and Brennan (2005) found evidence that speakers avoided syntactic ambiguities in deference to their addressees' ease of comprehension. With respect to referential forms, Englehardt, Bailey, and F. Ferreira (2006) show that whereas speakers are good at avoiding underspecified descriptions of entities in a scene, they are less good at avoiding overspecified descriptions of entities. In other words, people often produce descriptions with more information than is strictly necessary in a given situation, and listeners do not seem to judge these overdescriptions to be worse than more perspicuous descriptions. Continued work is needed to determine how more top-down considerations of addressee and conversational context and content are balanced against more bottom-up concerns about the visual characteristics and immediate salience of entities in conversations. Thorough summaries of recent psycholinguistic work in this area can be found in Konopka and Brown-Schmidt (2014) and Trueswell and Tanenhaus (2005).

3 Language Comprehension

Imagine that you want to learn about how a car engine works. So you find an engine that is running smoothly, and you open the hood. What would you learn by simply watching the engine running at its normal speed? The answer is likely not too much, simply because the engine is running very fast, and the connections between the various parts are too tightly linked to observe with the naked eye. In order to truly learn how each subcomponent of the engine operates and, in turn, interacts with other subcomponents, you would need to slow the engine down somehow – take out parts or introduce anomalies into the system somewhere, maybe even a literal monkey wrench. An analogous difficulty faces psycholinguists who study language comprehension: Comprehension is normally so fast and so smooth that it is nearly impossible to observe. As such, psycholinguists need a figurative monkey wrench to throw into the works. The most oft-tossed figurative wrench is ambiguity.

Ambiguity refers to situations in which a word (lexical ambiguity; see example (1) below, where *bank* is ambiguous) or sentence (structural ambiguity) can have two or more interpretations. There are also two subtypes of structural ambiguity, temporary and global. Temporary ambiguities (see example (3)) can be resolved by reading or listening to the rest of the sentence; global ambiguities cannot (see example (4)). Ambiguity is not the same thing as vagueness. Example (2) is simply vague because we do not know what Jack knew. Example (3) is temporarily ambiguous, and an initial, partial interpretation that "Jack knew the secret" must be revised to be interpreted correctly, specifically, that "Jack knew something about Jill, and that something is that the secret was bothering her." Note that example (3) does not technically state that Jack knew what the secret was. Example (4) is globally ambiguous because we do not know whether Jack was using a secret to blackmail the woman, or whether he was blackmailing a woman who had a secret (possibly wholly unrelated to the blackmail plot).

(1) Jack hid his ill-gotten gains in the bank by the river. (lexical ambiguity)

(2) Jack knew something. (vagueness)

(3) Jack knew the secret bothered Jill. (temporary structural ambiguity)

(4) Jack blackmailed the woman with a secret. (global structural ambiguity)

It is reported that around 40 percent of words in English are ambiguous (Traxler 2012), and a good bit of research has been conducted examining what happens in the mind when ambiguous words are encountered. The consensus of most of this work (e.g., Onifer and Swinney 1981; Seidenberg et al. 1982; Swinney 1979) is that all meanings of a word are accessed immediately, and very quickly thereafter (in less than 250 msec) the meanings that are unrelated to the immediate context are inhibited, leaving only the appropriate meaning behind.

When it comes to structural ambiguity, temporary structural ambiguity has proven to be particularly useful in slowing down the language comprehension system without forcing it to grind to a complete halt in the form of irreconcilable confusion (though see the discussion of interpretation below for some new findings). Sentences such as (3) have been termed "garden-path sentences" (Bever 1970) because they lead a reader or listener down a figurative garden path, which, at some point, is shown to be incorrect or misleading, and must be traced back to the error to recover from it. The initial error signal, and the subsequent recovery process, can be observed by psycholinguists in the form of disruptions to normal eye movement patterns during reading (e.g., Frazier and Rayner 1982; Pickering and Traxler 1998; Rayner 1998; Trueswell, Tanenhaus, and Garnsey 1994); eye movement patterns looking around scenes while listening to language (e.g., Tanenhaus et al. 1995); and patterns of brain waves called event-related potentials (ERPs) (e.g., Friederici 1998; Kutas, van Petten, and Kluender 2006) which are recorded via electroencephalography (EEG).

A critical factor in why we appear susceptible to these so-called garden-path disruptions is the "immediacy principle" proposed first by Just and Carpenter (1980). This principle states that, as much as possible, we derive interpretations from incoming language input immediately as it comes in, i.e., sequentially, word by word (Altmann and Kamide 1999; Marslen-Wilson and Tyler 1975; Tanenhaus 2007). We do not as a matter of course wait around to hear or read an entire sentence before we begin locating lexical meanings for words, constructing syntactic structure, and interpreting the meaning of the combined input. As a result, when we get to *the secret* in (3), we have built a syntactic structure in which *knew* takes a direct object, and *the secret* fills that slot nicely, so it gets automatically attached as the direct object. In general, the syntactic parser prefers to fill all possible argument positions in the simplest possible way, rather than closing off the verb phrase or positing more complex continuations for incoming material. Frazier (1978) proposed two syntactic parsing principles that captured these preferences, respectively: Late Closure and Minimal Attachment. These two principles have been the source and subject of hundreds of studies ever since, and, although they have been questioned and scrutinized on various grounds, they appear to be deeply entrenched preferences held by the human sentence processing system across languages.

Continuing with (3), the problem then arises when we get to the verb in the subordinate clause, *bothered*; this verb needs a subject, and *the secret* is the only option (in English). In order to assign subject status to *the secret*, it must be removed from object position of *knew*, and the resultant object role must be reassigned to the entire subordinate clause. As noted above, this process of "syntactic reanalysis" can be observed to some extent by looking at psychobehavioral measures such as eye movements during reading, which are tracked using infrared eye trackers. Frazier and Rayner (1982) provided the first eye-tracking evidence of syntactic reanalysis, documenting robust "garden-path effects" of re-reading, in the form of leftward eye movements (regressions), upon encountering disambiguating words (such as *bothered* in (3)).

The immediately preceding description of syntactic reanalysis assumes a certain theoretical position, however. Specifically, this theoretical position assumes that syntactic structures are built up not only incrementally, but also initially in the comprehension process and, importantly, independently of nonmorphosyntactic information sources (i.e., bottom-up). This

view of syntactic parsing in sentence comprehension motivated so-called serial, or two-stage models of syntactic parsing. These models are based conceptually on the view of the mind as having a compartmentalized, or modular, architecture (e.g., J. A. Fodor 1983), within which certain processes operate over only very specific types of information. This idea leads to a view of sentence parsing that (a) takes as primary the syntactic structure-building process, and (b) uses in this process only syntactic information at the first stage of parsing (e.g., Frazier 1978, 1987; Frazier and J. D. Fodor 1978; Rayner, Carlson, and Frazier 1983). The initial structure is pursued based solely on syntactic grounds, which, roughly speaking, consist of a few principles (including Late Closure and Minimal Attachment) specifying that the simplest possible structure be built first (e.g., Frazier 1978, 1987; Frazier and Flores D'Arcais 1989).

Under this very bottom-up view, other nonsyntactic information is not considered until after initial structure-building is performed, and then only if an error signal arises to alert the sentence processing mechanism that the constructed structure is inadequate to handle subsequent incoming material. Examples of these other types of material include semantic information such as plausibility (e.g., Ferreira and Clifton 1986) or probabilistic information such as whether a given verb (e.g., *expected*) is more likely to take a noun or a clause as a direct object (e.g., Garnsey et al. 1997; Trueswell, Tanenhaus, and Garnsey 1994; Trueswell, Tanenhaus, and Kello 1993). In other words, this sort of nonsyntactic, more top-down information does not have an effect under this two-stage view until a second stage in which reanalysis is performed (if necessary) and interpretations are generated.

An alternative, so-called single-stage view of sentence parsing is that all possible information sources – morphosyntactic, semantic, probabilistic, contextual, etc. – are computed in parallel. Within this class of "constraint-based" models, the relative influence of each of these various "constraints" on interpretation is calculated incrementally, and new values, or "weights," are assigned to each as more input comes in (e.g., MacDonald, Pearlmutter, and Seidenberg 1994; Tanenhaus et al. 1995; Trueswell, Tanenhaus, and Garnsey 1994), with bottom-up information and top-down information interacting throughout the process. Rather than the parser choosing one syntactic path to follow exclusively, all possible syntactic structures are considered in parallel from the start, ranked according to their likelihood of actually being correct based on the weights of the various constraints in play at the time. As the name of this class of models implies, no reanalysis is required in situations in which errors are recognized. Instead, processing slowdowns are attributed to the system having to reassign weightings to the various information sources, thereby inhibiting the activation, or availability, of the incorrect initial structure and boosting the activation of the correct structure.

A tremendous amount of ink has been spilled in an attempt to adjudicate between serial and parallel models of syntactic parsing, and exploring this vast literature even superficially would exhaust the space allotted for this chapter. An excellent summary of these sometimes contentious studies is provided by van Gompel and Pickering (2007), who point out that whereas some nonsyntactic information sources, such as discourse context, appear to guide the syntactic parser toward initial structures (e.g., Britt 1994; Liversedge et al. 1998; Spivey, Tanenhaus, Eberhard, and Sedivy 2002), semantic plausibility does not appear to (e.g., Clifton et al. 2003; Hoeks, Hendriks, Vonk, Brown, and Hagoort, 2006), and statistical probability (i.e., frequency) may exert such influence, but only under certain conditions or only weakly (e.g., Brysbaert and Mitchell 1996; Kennison 2001; Pickering, Traxler, and Crocker 2000).

A major reason why the rich, dense literature alluded to in the immediately preceding paragraph is not explored more deeply here is that the field of psycholinguistics has begun to move beyond this debate. More sophisticated, integrative models of the architecture of the system subserving language comprehension have recently been proposed, within which a certain degree of modular autonomy is maintained but parallel structure-building along the various levels or paths of linguistic representation is carried out (e.g., Jackendoff 2007, 2011), with frequency (lexical and structural) serving to moderate the computational speed of the various paths. Furthermore,

a resurgence of interest in the influence of top-down processing on language comprehension has resulted in computational models of sentence processing that emphasize the role of prediction in language comprehension (e.g., Clark 2013; Gibson, Bergen, and Piantadosi 2013; Hale 2003; Levy 2008; Van Petten and Luka 2012), which may go some way toward explaining why the effects of discourse context, mentioned above, appear to come online so rapidly.

It should be stressed, however, that some degree of prediction has long been acknowledged in the psycholinguistics literature. Consider the sentence in (5), in which the displaced noun phrase (NP) *the car* is associated with the object position of *wanted*. The displaced NP is called the "filler" and the position in the sentence with which the filler is associated is called the "gap."

(5) That's the car that Sophie wanted [gap] for her birthday.

J. D. Fodor (1978, 1989) proposed the active filler hypothesis, to explain why researchers observed that *car* is more active in people's minds at the gap site than at other places in the sentence, and more active than other nouns (e.g., Nicol and Pickering 1993; Nicol and Swinney 1989). The active filler hypothesis also explains why people experience processing difficulty at the potential gap site in (6) once the actual gap site is encountered (cf. Phillips 2006).

(6) That's the car that Sophie wanted [potential gap] her parents to buy [actual gap] for her birthday.

The active filler hypothesis says that the syntactic parser seeks to associate all NPs in non-case-assigning positions with case-assigning positions as soon as possible. In essence, the gap is predicted somewhere downstream as soon as *the car* is encountered. Note, though, that this prediction is tightly linked to syntactic well-formedness requirements – case and thematic role assignment – not strictly because cars are things that people want or buy. Under this syntax-centric view, but not a constraint-based view, we should observe an equal amount of processing difficulty at *wanted* in (6) and in (7) (a comparison which, I believe, has not been directly tested).

(7) It was the typhoid that Sophie wanted [potential gap] to cure [actual gap] quickly.

Predictions about upcoming input based also on nonsyntactic information are receiving more and more attention, often couched in terms of "surprisal" at unexpected incoming material (Hale 2003; Levy 2008). Researchers have speculated for some time that distributional probabilities within language may guide structure-building during comprehension (Elman 1991; Kimball 1975; MacDonald, Pearlmutter, and Seidenberg 1994), but recent advances in "big data" analysis (e.g., Baayen 2008) and computational modeling now allow these probabilistic linguistic tendencies to be explicitly tested as predictors of comprehension (and production). Yet the breadth and precision of predictions that are made during language processing must logically be limited, as it does not seem plausible that we entertain every possible continuation of every sentence at every juncture. As mentioned previously, though, under some circumstances quite specific predictions appear to be made, including the letter order of upcoming words (Luke and Christianson 2012) despite the fact that letter order is often encoded with some degree of flexibility (e.g., Perea and Lupker 2003, 2004). Regarding syntax, Staub and Clifton (2006) show that when readers encounter *either*, they predict the existence of an upcoming *or*. One famous example of evidence that semantic predictions are made, and that disruption of some sort is encountered when those predictions are not borne out, is the N400, which is a brain wave pattern observed in EEG recordings as people read sentences such as *Mary spread the warm bread with socks/butter*. Using this event-related potentials (ERP) method, the N400 has been shown to be reliably larger when the unexpected critical word (*socks*) is encountered compared to a predictable word (*butter*) (e.g., Kutas and Hillyard 1980, 1983).

In all of these cases, predictions about upcoming input are based on previous material (context). Then when the new input is encountered, the degree to which it satisfies predictions can

only be determined by checking it against the previous material in some way. As a concrete example, even *socks* in the Kutas and Hillyard sentence above would not be anomalous if we could not remember that we had just read *spread* and *bread*. So the usefulness of prediction in language processing is limited in one direction by countless possible continuations, even in constrained contexts (butter? jam? jelly? honey? Nutella? liverwurst?) (Jackendoff, 2002), and in the other direction by the relatively small capacity of working memory and fallibility of retrieval processes. The limitations that memory places on language processing were demonstrated by Chomsky (1956), who showed that an infinite number of center-embeddings are grammatically allowable in human language (cf. Legate, Pesetsky, and Yang 2014). Yet, despite their grammaticality, more than one center-embedding (e.g., 8–9, from Pulman 1986) becomes nearly impossible to interpret (Chomsky and Miller 1963; Miller and Isard 1963). More recently, several language processing models have provided accounts to explain a number of asymmetries in terms of working memory and/or recall limitations. Perhaps most well known among these asymmetries is the relative difficulty of processing object-relative clauses (11) compared to subject-relative clauses (12).

(8) The woman the man the girl met loved died. (double center-embedding)

(9) The woman the man the girl the dog bit met loved died. (triple center-embedding)

(10) The woman somebody I met loved died.

(11) The horse that the cow frightened ran outside. (object-relative clause)

(12) The horse that frightened the cow ran outside. (subject-relative clause)

(13) The horse that the radio frightened ran outside.

In the interest of space, I highlight only two of these recent memory-based models of language processing. The first was initially introduced by Gibson (1998) as the syntactic prediction locality theory, and later refined by Gibson (2000) as the dependency locality theory. The key insight in both of these theories is that language processing imposes two types of cognitive costs, storage costs and integration costs. Storage costs accrue as elements are held longer in memory; in particular, costs accrue for arguments and their heads, and the longer an argument must be held in memory before being assigned case/thematic role from the head, the greater the cost. Thus, the object-relative clause is harder than the subject-relative clause because the extracted object (*horse*) has to be held in memory while the subject of the relative clause *(cow*) is assigned its case and thematic role by the verb, and multiple center-embedded structures require progressively more storage. Integration costs accrue as more elements need to be integrated at a given point. Again, when the relative clause verb (*frightened*) is encountered in the object-relative clause, both *horse* and *cow* must be integrated, whereas in the subject-relative clause, the extracted subject (*horse*) is already integrated before *cow* is encountered. The latter dependency locality theory was proposed to account for data observed by Warren and Gibson (2002), who investigated the fact that third-person nouns and pronouns feel intuitively to place more burden on memory than first- or second-person pronouns (compare (8) to (10)). Both versions hinge on structurally defined distance between syntactic dependencies to explain what are, as a class, called *locality effects* – observations that dependent elements are generally processed more slowly when they are separated further from one another.

A different approach to working memory and language processing has been proposed by Van Dyke and Lewis (2003), and Lewis and Vasishth (2005) and Van Dyke (2007), which also accounts for *antilocality effects* – the counterintuitive finding that, under some circumstances, increasing the distance between syntactically dependent elements actually speeds the processing of the latter element (e.g., Vasishth and Lewis 2006). This family of processing theories is collectively known as retrieval-activation accounts or similarity-based interference accounts (or a few other related monikers), and rely on the general principles of activation and decay that have long been used to explain forgetting in memory research (cf. Altmann and Schunn 2012; Anderson et al. 2004; Anderson and Lebiere 1998). The basic idea is that as linguistic elements arrive in the input,

their concepts and features (on various levels of analysis) are activated. This activation decays over time, which would be the only problem if language did not contain long-distance dependencies (such as displaced elements). But language does contain long-distance dependencies, so sometimes intervening elements are encountered that share certain features with earlier activated elements. When those earlier activated elements need to be reactivated to be integrated into the structure being built, they are interfered with by more recently encountered elements. Importantly, this interference is proportional to the amount of featural overlap the two elements share. To take a concrete example, (11) is predicted to be harder than (12) because *horse* needs to be reactivated as the object of *frightened*, but *cow* intervenes. Because cows and horses share a great number of syntactic and semantic features, *cow* triggers interference in reactivating *horse*. Under this account, (11) is predicted to be more difficult than (13): whereas both cows and horses can frighten and be frightened (and can both run), radios cannot be frightened (nor can they run), and thus *radio* should interfere less with the reactivation of *horse* than *cow* does. Interestingly, Staub (2010) presents eye movement data from normal silent reading of object-relative clauses that suggest independent effects of predictions (i.e., surprisal; Levy 2008) and memory retrieval difficulty (see also Vasishth and Drenhaus 2011). These data suggest that both processes contribute to the difficulty of processing object-relative clauses, and, by assumption, other difficult structures.

Federmeier (2007) also highlights the coexistence of predictive, top-down processes and integrative, bottom-up processing in the Production Affects Reception in Left Only (PARLO) framework. This framework is predicated on experimental results of ERP, functional magnetic resonance imagery (fMRI), and visual half-field methods, in which words are presented in only one visual field or the other to assess the relative contribution of the left and right brain hemispheres. Federmeier argues the combined data suggest that the left hemisphere operates over more abstract representations of the input, and is also involved in more top-down prediction of upcoming material. The right hemisphere appears to operate over more veridical, concrete representations, and proceeds in a more bottom-up, integrative fashion. Understanding precisely how these types of processing are coordinated across tasks and across comprehenders who differ in their cognitive profiles requires further exploration using combined behavioral and neurophysiological/neuroimaging techniques.

As suggested by the immediately preceding discussion, a greater understanding of the role of general cognitive processes and constraints, such as working memory capacity, verbal fluency, and language/text exposure (e.g., Daneman and Carpenter 1980; Engle 2010; Just and Carpenter 1992; Payne et al. 2014), is emerging (though see, e.g., Caplan and Waters 1999, 2002; Caplan, Waters, and DeDe 2007, for arguments that language processing is not impacted by decrements in more general cognitive abilities). This line of research highlights the role of individual differences in how both bottom-up information and top-down information get integrated during language comprehension (e.g., Christianson et al. 2006; Kim and Osterhout 2005; MacDonald, Just, and Carpenter 1992; Swets et al. 2007; Tanner 2013; Traxler 2009; Traxler et al. 2005). Let us now touch upon these topics in turn as we consider how far the above-referenced work on language *comprehension* takes us in our ultimate understanding of language *interpretation*.

A great deal of psycholinguistic work often uses the terms "parsing" and "processing" interchangeably, and sometimes tosses "comprehension" into the terminological blender, mixing all three together. It would be helpful to more carefully delimit these terms, using "parsing" to refer exclusively to morphosyntactic structure-building and "processing" to refer to the integration of all available information. "Comprehension" should be reserved to refer to the way in which readers and listeners mentally represent the actual content of the input. Finally, "interpretation" can be used to refer to the integration of the comprehended content (or the final mental representation) of the input into the larger context of the discourse and the interlocutors' knowledge of both the immediate situation and the world. Chomsky has taken a rather pessimistic view of the ultimate success in understanding language "interpretation," as defined here. He argues:

> …[It is a problem] to construct an "interpreter" which includes the parser as a component along with all other capacities of the mind – whatever they may be – and accepts nonlinguistic as well as linguistic inputs. This interpreter, presented with an utterance and a situation, assigns some interpretation to what is being said by the person in the situation. The study of communication in the actual world of experience is the study of the interpreter, but this is not a topic for empirical inquiry, for the usual reasons: there is no such thing as the study of everything. (Chomsky 2000, p. 69)

Chomsky is correct in pointing out that it is impossible to empirically control for all of a person's past experiences, summed world knowledge, representation of the mind-state of other interlocutors, and present emotional state to determine, unequivocally, how a given person is at the very moment interpreting a given sentence. What Chomsky implies, therefore, is that only the parser, and its work of parsing input (as defined above), is sufficiently narrow to be empirically studied. On the other hand, it can be forcefully argued that shying away from interpretation, or at least the failure to probe explicit comprehension, fails to provide a clear picture not only of comprehension, but even of the parser itself. Lyn Frazier (1998), in an excellent review of psycholinguistic advances through the end of the twentieth century, argues that "deciding between alternative accounts of syntactic parsing may depend increasingly on an explicit account of how and when the constituents of a sentence are interpreted" (p. 137). Furthermore, says Frazier, developing theories of interpretation can be considered to be "the central task of psycholinguists in the future" (p. 138). What appears to be a rather substantial hurdle on the way to addressing the "central task" of illuminating interpretation is the widespread, generally standard assumption that comprehension is reflected straightforwardly by implicit measures of online syntactic parsing.

As an example of implicit measures of online syntactic parsing, let us consider eye movement patterns. An enormous amount of psycholinguistic research has been carried out examining the way people's eyes move as they read text (cf. Rayner 1998, 2009; Staub and Rayner 2007). As people read, their eyes do not scan smoothly across the text; rather, the eyes proceed in a series of stops (fixations) and jumps (saccades). During the fixations (which average around 200–250 msec), information is processed – information about the material within the range of the current fixation, and some limited information about upcoming material (cf. Reichle, Rayner, and Pollatsek 2003) – and during the saccades (which generally last less than 30 msec), readers are essentially blind. It has been well established that fixation durations are longer when processing more difficult material. Similarly, forward saccades are shorter and more frequent as the difficulty of the material increases, and there are more backward saccades (regressions) when reading bogs down, especially in connection to error signals related to the syntactic parse (e.g., in the context of garden-path sentences) (Rayner et al. 2012).

This link between eye movements and the immediate difficulty of constructing the syntactic parse is so well established, in fact, that it is tempting to assume it to be relatively trivial to connect eye movements with comprehension. As an example of this assumption, consider the following quote from Staub and Rayner (2007: 327): "Because eye movements are a natural part of the reading process, secondary tasks are not needed to make inferences about reading comprehension." Certainly such inferences can be made, but how accurate are they? Consider the garden-path sentence in (14a) and its unambiguous counterpart in (14b) (from Christianson et al. 2001; Ferreira, Christianson, and Hollingworth 2001), which are termed DO/S ambiguities because a noun phrase (e.g., *the baby*) is initially parsed as the **d**irect **o**bject of the subordinate verb (e.g., *dressed*) and subsequently must be reanalyzed as the **s**ubject of the main clause verb (e.g., *spit up*).

(14) a. While Anna dressed the baby that was cute and cuddly spit up on the bed.
b. While Anna dressed, the baby that was cute and cuddly spit up on the bed.

A number of eye tracking studies have demonstrated that readers have trouble with sentences such as (14a) compared to sentences such as (14b) (e.g., Pickering and Traxler 1998; Slattery et al.

2013). Specifically, fixation durations on the main verb (in this case *spit up*) are inflated in (14a) compared to (14b), and readers regress more in (14a), going back to reread the sentence. Following Frazier and Rayner (1982), both of these signs of disruption are interpreted as signaling that the reader has noticed a problem with the ongoing parse, and the rereading is conducted to reanalyze the parse. Crucially, this revision has been assumed to result in a revised interpretation as well. Yet rereading might not be any more successful than initial reading, or it might only be partially successful, or it might signal much deeper comprehension. In fact, we do not yet know for certain how these online eye movement patterns map onto actual comprehension, defined above as the final semantic representation derived from the input.

We have reason to believe, however, that comprehension is susceptible to a number of influences, which stem from the overarching goal of language processing: to derive a coherent, plausible interpretation (Ferreira and Patson 2007). Evidence that people misinterpret sentences more often and far more systematically than previously suspected comes from many sources. Phillips, Wagers, and Lau (2011) provide an excellent summary of various *grammatical illusions*, in which ungrammaticalities are either produced or not noticed (e.g. (15); see, e.g., Bock and Miller 1991; Eberhard, Cutting, and Bock 2005; Solomon and Pearlmutter 2004).

(15) The key to the cabinets are on the table.

The fact that ungrammaticalities in grammatical illusions are not noticed is curious, given that other syntactic constraints appear to trigger processing difficulty immediately upon encounter, even in young children. An example would be Principle C of binding theory (Chomsky 1981), as seen in (16) (from Crain and McKee 1985), which places constraints on cataphoric coreference.

(16) a. While he_i was eating pizza, Ninja $Turtle_i$ was dancing.
b. *He_i was eating pizza while Ninja $Turtle_i$ was dancing.

Consider again the garden-path sentence in (14a). It has been shown a number of times now that people do not fully recover from the initial misinterpretation that *the baby* is being dressed (Christianson et al. 2001, 2006; Patson et al. 2009; Slattery et al. 2013; van Gompel et al. 2006), despite the fact that they both notice that something is wrong at the point of disambiguation (*spit up*) and make regressive eye movements to re-read these sentences. In fact, Christianson and colleagues have asked participants explicitly in various experiments whether the baby was being dressed or whether the baby was spitting up, and large proportions of participants reported believing both to be true, despite the fact that this interpretation should be ruled out by the syntax. Christianson et al. (2006) additionally found that older adults with less working memory capacity were only at chance in believing that *Anna* was dressing herself (as the syntax stipulates). As such, this sort of misinterpretation appears to be a variety of grammatical illusion, in that despite disruption at the point of disambiguation, people apparently forge ahead with a plausible yet structurally illicit interpretation.

One possibility for both grammatical illusions and more general misinterpretation effects is that people are uncertain about the previous material that they have encountered (e.g., Levy et al. 2009). Another possibility, discussed above, is that the language processor balances structure-building (parsing) with meaning-making (processing) en route to the ultimate goal, interpretation. And, in cases where the structural and meaning-based ouputs do not converge, one is chosen over the other without concomitant revision of the other output. The full range and implications of these possibilities have not been thoroughly explored as of yet, but they can be classified as "underspecification" or "good enough" accounts (cf., Christianson et al. 2001, 2006; Christianson, Luke, and F. Ferreira 2010; F. Ferreira 2003; F. Ferreira, Bailey, and Ferraro 2002; F. Ferreira and Patson 2007), in that the ultimate interpretation is based on an analysis of the input that is somehow less than certain or complete. This view of language processing is related to the concept of "satisficing" (Simon 1947) in the broader realm of cognitive science.

Interestingly, individual differences in the way people process input that pits syntax against semantics have been observed by a number of researchers using ERPs to monitor brain activity while reading sentences such as (17) from Kim and Osterhout (2005) (e.g., Kim and Osterhout 2005; Kolk et al. 2003; van Herten, Kolk, and Chwilla 2005).

(17) The hungry meal was devouring the kids.

Space limitations prevent a full discussion of these individual differences. Suffice it to say that such sentences elicit from some people in these studies a so-called N400 effect, traditionally associated with semantic anomalies (see discussion above), while from other people, a so-called P600 effect is elicited. The P600 is traditionally associated with syntactic anomalies or errors (Friederici 1998). It thus appears that processing mechanisms might address anomalies either semantically or syntactically, as suggested in models proposed by Kuperberg (2007; Kuperberg et al. 2006) and Bornkessel and Schlesewsky (2006; Schlesewsky and Bornkessel 2006) based on ERP data, and as suggested by behavioral data by, e.g., F. Ferreira (2003) and Christianson, Luke, and Ferreira (2010). Moreover, the strategy that a given individual's language processor applies to any particular challenging input may well depend on the cognitive resources, like working memory and cognitive control, that are available either in general or in response to a particular task (cf. Christianson et al. 2006; Christianson and Luke 2011; F. Ferreira and Patson 2007; Hussey and Novick 2012; Novick et al. 2014; Swets et al. 2007; Tanner 2013; Ullman 2004). The precise mechanisms underlying individual differences in language processing are yet to be determined. It has not yet even been established whether the differences between individuals are stable within individuals over time. Nevertheless, these differences promise to provide insight into the various tensions we have discussed: top-down vs. bottom-up processing, prediction vs. memory retrieval, serial vs. parallel processing, and structure vs. meaning.

One might imagine that recent advances in brain-imaging technology would help answer questions about individual differences and the connection between language and other cognitive capacities. Although technologies such as functional magnetic resonance imagery (fMRI) are extremely promising, and have yielded intriguing results, continuing disagreements about methodologies and data analysis and interpretation make it difficult to draw firm conclusions just yet. Let us briefly consider two cases in point.

First, fMRI has been used to investigate the field-defining question of how language interacts with more general cognitive functions. Paul Broca (1861) first identified the left inferior frontal gyrus (LIFG) as being involved in language, specifically in syntactic processing. A robust debate still rages, however, about whether this and associated brain regions have language-specific functions (Fedorenko, Behr, and Kanwisher 2011; Fedorenko and Thompson-Schill 2014) or more domain-general functions, such as cognitive control (January, Trueswell, and Thompson-Schill 2009; Novick, Trueswell, and Thompson-Schill 2005) or phonological short-term memory (Rogalsky and Hickock 2011), that are recruited for language processing. The basic approach taken in these investigations is to administer to participants linguistic and nonlinguistic tasks of various sorts while participants are in an fMRI magnet. Then brain regions traditionally associated with linguistic processing are examined for activation during the various tasks. When areas of activation overlap across linguistic and nonlinguistic tasks, it is inferred that the area is performing computations that are involved in, but not unique to, language processing. As noted, however, differences in opinion about the significance and type of overlap, and the components of any given task, have proved to be stumbling blocks on the path to broad consensus.

Another intriguing recent finding related to how language interacts with the rest of the brain is that the left premotor cortex, which is involved in generating sensory-motor representations of physical actions, shows similar patterns of activation when both observing physical actions such as biting a peach and also reading phrases such as *biting the peach* (Aziz-Zadeh et al. 2006). This finding is interpreted within the theory of embodied cognition (e.g., Barsalou 1999; Glenberg and Kaschak 2002; Zwaan 2004), which posits that language comprehension (and various

other aspects of cognition) is grounded in our kinesthetic experience with the world (see also the much older motor theory of speech perception, e.g., Liberman and Mattingly 1985). The basic idea is that a class of neurons called mirror neurons, found in the premotor cortex of primates, fire when monkeys (and humans, according to the theory) perform a task or see it performed. In humans, the proposal is that these mirror neurons also fire when hearing or reading about physical actions, and we derive the meaning of these words from this shared firing. In other words, we know what *kick* means because we know how to kick. Of course, human semantics encompasses far more types of meaning than physical actions, and researchers have sought to connect mirror neurons in the sensorimotor and limbic systems to the semantics of perception and emotion (Pulvermüller 2013). Here, too, however, we are yet far from consensus, as other researchers dispute the centrality of mirror neurons in language processing (e.g., Lotto, Hickock, and Holt 2009).

4 Developmental Psycholinguistics

One of the most basic, defining questions in both linguistics and psycholinguistics is how children come to acquire and use language. Accordingly, there is a long history of empirical investigation into how children process language at all levels of analysis. A primary concern of this research has been to determine whether the language processing of young children differs from that of adults, and, if so, how. The logic behind this question is as follows: By removing the vagaries of experience, we can reduce individual differences deriving from that experience, and thus where children and adults show similar processing, we can attribute it to innate or hardwired underlying mechanisms. See Chapter 19 for more detailed discussion.

Of course, it is not a trivial task to assess the language processing and comprehension of children, ranging in age from early elementary school down to newborn. A number of clever methods have been devised to elicit behavioral data from which internal processing states can be inferred (e.g., Hoff 2012; McDaniel, McKee, and Cairns 1996). These include the *high amplitude sucking* paradigm, in which infants are given a non-nutritive nipple with a pressure transducer inside to suck on as they are played recordings of various linguistic input. As the recordings become tedious, the sucking rate slows; when the recording changes, the sucking intensity might or might not also increase. If it does, it is taken as evidence that the infant was sensitive to the change. A similar idea underlies the *conditioned head turn procedure*, which is used with older infants who have better control of their gaze directionality. Audio recordings are played in concert with rewarding visual stimuli, which appear when the audio changes in some theoretically interesting way. Infants begin to associate these pairings and anticipatory head turns are taken as evidence of recognition of the given change. Similarly, the *intermodal preferential looking paradigm* and *looking-while-listening procedure* both pair pictures or video recordings depicting two actions (say, Big Bird tickling Cookie Monster in one video, and Cookie Monster tickling Big Bird in the other video) and an audio recording such as *Big Bird is tickling Cookie Monster. Where is Big Bird tickling Cookie Monster?* (Hirsch-Pasek and Golinkoff 1996). The infant's or young child's gaze is monitored to assess the speed with which the child turns toward the picture or video matching the audio as well as how long the child looks at the correct video.

The reader is encouraged to refer to more comprehensive compendiums of developmental psycholinguistics and language development research (e.g., Hoff 2012, 2014; Jusczyk 1999; McDaniel, McKee, and Cairns 1996; Seidl and Cristiá 2012; Trott, Dobbinson, and Griffiths 2004; Werker, Yueng, and Yoshida 2012), as we are limited here to but a small sample of this fascinating work. For example, newborns between 0 and 5 days old have been found to distinguish between the sounds of different languages (Byers-Heinlein, Burns, and Werker 2010), including the two languages spoken by their bilingual mothers. At just two months of age, English-learning infants can already distinguish between allophonic variations in minimal pairs such as the /t/ in "night rate" and "nitrate" (Hohne and Jusczyk 1994). By about the age of ten months, infants appear to

have developed preferences for the phonemic contrasts and allophonic rules of their native (or familiar) language compared to those of other languages, whereas, prior to that age, they appear to entertain all contrasts and allophonic rules allowed in human language (e.g., Jusczyk et al. 1993; Seidl et al. 2009; Werker et al. 2012).

A central focus of much (but not all) of the work examining how very young children process language has been to determine the mechanisms by which children come to have phonological, morphological, lexical, and syntactic categories or rules. In a groundbreaking study Saffran, Aslin, and Newport (1996) demonstrated that eight-month-olds appear to use transitional probabilities between "words" in continuous speech. Saffran et al. played two-minute recordings of three-syllable pseudowords such as (18) (shown with "word" transitions demarcated).

(18) tupiro|golabu|bidaku|padoti|bidaku|golabu|bidaku|tupiro…

After a short pause, they then played a second recording of either the same syllables combined in the same ways as in the first recording (e.g., *tu* always followed by *pi* and *pi* always followed by *ro*), or combined in different ways. They found that infants who were played the recording containing the new combinations listened longer than those who were played the familiar recording. The inference was that the new combinations captured the infants' attention as they worked on learning the new transitional probabilities between the syllables.

This result and many like it – results that have shown that children are able to utilize probabilistic regularities in the input in their language learning – have been taken by some as evidence against the innateness of language and Universal Grammar (e.g., Chomsky 1959, 1975); however, as argued by Yang (2004), statistical learning like this is not at all inconsistent with innateness or Universal Grammar. Furthermore, it seems that computer models of purely statistical learning are significantly improved when bolstered by other learning principles or strategies (e.g., McMurray, Aslin, and Toscano 2009; Swingley 2009; Werker et al. 2012; Yang 2004) that are available to infants.

Infants also use syntactically sophisticated probabilistic information to learn the meanings of new words. For example, infants use the subcategorization frames of verbs (e.g., whether they are transitive or intransitive) to learn their meanings (*syntactic bootstrapping*, e.g., Gleitman 1990; Gleitman et al. 2005; Naigles 1990). Recent studies have shown that children between 19 and 21 months old can use the number of nouns in a sentence to establish the subcategorization frame, and consequently interpretation, of novel verbs (Yuan, Fisher, and Snedeker 2012). The mechanisms underlying infants' ability to map words on to scenes and then attach meaning on to those words are still a matter of considerable debate. Although it is possible that infants store vast statistical information about the co-occurrence of words and entities in the world, it is also the case that the world is a very busy, messy place, and the accuracy of word–world mappings (and associated word learning) decreases considerably once scenes become more complex (Scott and Fisher 2012).

EEG and certain brain imaging techniques (magnetoencephalography, or MEG) can also be used on relatively young infants. Space limitations preclude even a cursory discussion of these studies, however. Suffice it to say here that these methods promise continuing insights into the sophisticated analysis of incoming linguistic input that is performed by even prelingual infants. For example, Kooijman, Hagoort, and Cutler (2005) show that infants quickly and efficiently extract familiar words from connected speech, even in the absence of any semantic association for these words (i.e., words the infants have heard but do not yet know the meanings of). Once some semantic associations are established, however, it appears that 12 to 18-month-old infants' brains are activated differently by familiar words compared to novel words and nonwords (Travis et al. 2011). Moreover, the activation occurs in infants in the same areas of the brain and along similar time courses as in adults. Finally, when comparing the syntactic processing of children to that of adults, what appears to differ between the two groups is the development of cognitive control, rather than access to morphosyntactic principles. Across the few typologically distinct languages

that have been examined, it seems to be the case that four- to five-year-old children have more difficulty than adults using later-arriving information (be it structural or lexical information) to recover from initial misinterpretations (Choi and Trueswell 2010; Trueswell and Gleitman 2007; Trueswell et al. 1999). This series of results once again points to the importance of individual differences at the interfaces of language and other cognitive capacities.

5 Applied Psycholinguistics

Although the term "applied psycholinguistics" can refer to several fields, we limit the discussion here to second/foreign language processing. It is only recently that a truly psycholinguistic theory has been developed for language processing by non-native speakers. A major reason for this is that second language acquisition research has historically been concerned with the practical issue of improving foreign language pedagogy or with the broader theoretical issues of whether native and non-native languages were represented the same way in the mind and whether non-native language learning was constrained by Universal Grammar (e.g., Larsen-Freeman, Long, and Jiang 1991; White 2003). In 2006, Clahsen and Felser (2006a, 2006b, 2006c) proposed the *Shallow Structure Hypothesis* (SSH), which effectively reoriented second language researchers toward a more psycholinguistic approach.

What Clahsen and Felser proposed in the SSH was this: It has been observed that even advanced non-native speakers produce certain perseverative errors when they speak or write in their second languages (L2), and furthermore, certain grammatical errors are often not noticed when listening or reading in an L2. This observation about real-time (online) processing stands in stark contrast to the fact that in offline tests of explicit grammatical knowledge, these same L2 speakers often display full conscious command of these very same grammatical principles. The SSH says that the problem is in the real-time processing of the L2, rather than in the underlying knowledge of the L2 (which had been the focus of much of the past decades of L2 research). Clahsen and Felser proposed that non-native speakers are generally incapable of applying the grammatical principles that they know in real-time during processing.

The SSH has spurred a tremendous amount of research testing its claims. Some has provided evidence in support of the SSH and some against (e.g., Clahsen et al. 2010; Felser and Roberts 2007; Felser et al. 2003; Foote 2011; Frenck-Mestre 2005; Havik et al. 2009; Hopp 2006; Jiang 2004, 2007; Lim and Christianson 2013a, 2013b, 2014; Marinis et al. 2005; McDonald 2006; Papadopoulou and Clahsen 2003; Tokowicz and Warren 2010; Williams 2006). As one might imagine, the jury is still out as to the precise strengths and weaknesses of the SSH. The crux of the debate at present is whether L2 processing is qualitatively different from L1 processing, or quantitatively different. Additionally, whatever the answer turns out to be, there is a further question of whether or not it must necessarily differ. The distinction between parsing and processing discussed above might be useful in evaluating the results, and in designing future studies. Is the question whether non-native speakers *parse* L2 input differently from native speakers? Or is the question whether they *process* L2 input differently? If parsing is the issue, then it means that the grammatical constraints of the L2 are not being applied adequately in real-time. But if the grammatical constraints of the L2 are being adhered to and applied appropriately to the input, then processing is the issue, and the observed errors derive from inadequate integration of syntactic and semantic/pragmatic information. Some recent research seems to show evidence of a full morphosyntactic parse coexisting with evidence of semantic misinterpretation (e.g., Lim and Christianson, 2013a, 2013b), but much more work is required to evaluate these theoretical possibilities.

Applied psycholinguistics is a field wrought with complications. Unlike developmental psycholinguistics, in which research subjects (infants and children) are considered to be relatively homogeneous and limited in their experience, adult L2 speakers vary in myriad ways, including age, education, first language (L1), literacy level (in L2 and L1), L2 proficiency level, working

memory capacity and other cognitive capacities, etc., etc. Perhaps this is why earlier applied psycholinguistic research rarely utilized the same methods as psycholinguistic research with monolinguals (e.g., eye-tracking, ERP, brain imaging). Recent adoption of linear effects modeling in psycholinguistic data analysis (e.g., Baayen 2008; Baayen, Davidson, and Bates 2008; Jaeger 2008), however, allows for inclusion of these many factors in experimental designs and analyses, accounting for even widely divergent individual differences, which may turn out to be more influential than crosslinguistic differences (Kim and Christianson 2013; Tanner 2013).

Given that most of the world's population is at least bilingual (Grosjean 2010), it is a welcome development in psycholinguistics that bilinguals and adult language learners are no longer treated as deviant or special cases in comparison to monolinguals. We have barely touched upon a couple of important issues in this area of psycholinguistics here; the interested reader is encouraged to refer to Grosjean and Li (2012) and Grosjean (2008) for reviews of the multifaceted literature. As we learn more about individual differences even among monolingual speakers, the interfaces between language and various other cognitive functions are coming more and more to the forefront of psycholinguistic research. Whereas once variability in language processing was regarded largely as noise in the data (but cf. Just and Carpenter 1992), now this variability is becoming appreciated as valuable indication of the exogenous and endogenous constraints placed on language comprehension and production in real-time language use. And conversely, we are beginning to more fully appreciate how language processing also shapes and hones the cognitive capacities with which it interfaces (e.g., Bialystok and Barac 2012; Bialystok, Craik, and Freedman 2007; Fausey and Boroditsky 2010, 2011; Senay, Albarracín, and Noguchi 2010; Zwaan and Radvansky 1998). It is a clearer understanding of these interfaces that promises to dispel the "tensions" considered in this chapter, as well as to alleviate tensions between the related fields of linguistics and psychology.

REFERENCES

Altmann, E. M. and Schunn, C. D. (2012). Decay versus interference: a new look at an old interaction. *Psychological Science*, 23, 1435–7.

Altmann, G. T. M. and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73, 247–64.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of mind. *Psychological Review*, 111, 1036–60.

Anderson, J. R. and Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Erlbaum.

Arnold, J. E., Wasow, T., Asudeh, A., and Alrenga, P. (2004). Avoiding attachment ambiguities: the role of constituent ordering. *Journal of Memory and Language*, 51, 55–70.

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., and Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, 16, 18, 1818–23.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 4, 390–412.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*, 637–60.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (ed.), *Cognition and the Development of Language*. New York: Wiley.

Bialystok, E. and Barac, R. (2012). Emerging bilingualism: dissociating advantages for metalinguistic awareness and executive control. *Cognition*, 122, 67–73.

Bialystok, E., Craik, F. I., and Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, 45, 459–64.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–87.

Bock, J. K. and Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23, 45–93.

Bock, K. and Griffin, Z. M. (2000). The persistence of structural priming: transient activation or

implicit learning? *Journal of Experimental Psychology: General*, 129, 2, 177.
Bock, K. and Warren, R. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21, 1, 47–67.
Bornkessel, I. and Schlesewsky, M. (2006). The extended argument dependency model: a eurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113, 787.
Brennan, S. E., and Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1482–93.
Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, 33, 251–83.
Broca, P. P. (1861). Loss of speech, chronic softening and partial destruction of the anterior left lobe of the brain, translated by Christopher D. Green. *Bulletin de la Société Anthropologique*, 2, 235–8. Available online at http://psychclassics.yorku.ca.
Brown-Schmidt, S., Campana, E., and Tanenhaus, M. K. (2005). Real-time reference resolution by naive participants during a task-based unscripted conversation. In J. C. Trueswell and M. K. Tanenhaus (eds), *Approaches to Studying World-Situated Language Use: Bridging the Language as Product and Language as Action Traditions* (pp. 153–72). Cambridge, MA: MIT Press.
Brown-Schmidt, S. and Tanenhaus, M. K. (2006). Watching the eyes when talking about size: an investigation of message formulation and utterance planning. *Journal of Memory and Language*, 54, 592–609.
Brysbaert, M. and Mitchell, D. C. (1996). Modifier attachment in sentence parsing: evidence from Dutch. *Quarterly Journal of Experimental Psychology*, 49A, 664–95.
Byers-Heinlein, K., Burns, T. C., and Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science*, 21, 343–8.
Caplan, D. and Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22, 77–94.
Caplan, D. and Waters, G. S. (2002). Working memory and connectionist models of parsing: A response to MacDonald and Christiansen (2002). *Psychological Review, 109,* 66–74.
Caplan, D., Waters, G., and DeDe, G. (2007). Specialized working memory for language comprehension. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, and J. N. Towse (eds), *Variation in Working Memory* (pp. 272–302). Oxford: Oxford University Press.
Chang, F., Dell, G. S., and Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113, 2, 234.
Chen, B., Jia, Y., Wang, Z., Dunlap, S., and Shin, J. A. (2013). Is word-order similarity necessary for cross-linguistic structural priming? *Second Language Research*, 29, 4, 375–89.
Choi, Y. and Trueswell, J. C. (2010). Children's (in) ability to recover from garden paths in a verb-final language: evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, 106, 41–61.
Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, 2, 113–24.
Chomsky, N. (1959). A note on phrase structure grammars. *Information and Control*, 2, 393–5.
Chomsky, N. (1975). *The Logical Structure of Linguistic Theory*. New York: Plenum Press.
Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
Chomsky, N. (2000). Minimalist inquiries: the framework. In R. Martin, D. Michaels, and J. Uriagereka (eds), *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik* (pp. 89–156). Cambridge, MA: MIT Press.
Chomsky, N. and Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, and E. Galanter (eds), *Handbook of mathematical psychology* (pp. 269–321). New York: Wiley.
Christianson, K. and Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition*, 98, 105–35.
Christianson, K., Hollingworth, A., Halliwell, J. F. and Ferreira, F. (2001). Thematic-roles assigned along the garden path linger. *Cognitive Psychology*, 42, 368–407.
Christianson, K. and Luke, S. G. (2011). Context strengthens initial misinterpretations of text. *Scientific Studies of Reading*, 15, 136–66.
Christianson, K., Luke, S. G., and Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 538–44.
Christianson, K., Williams, C. C., Zacks, R. T., and Ferreira, F. (2006). Misinterpretations of garden-path sentences by older and younger adults. *Discourse Processes*, 42, 205–38.
Clahsen, H. and Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, 27, 3–42.
Clahsen, H. and Felser, C. (2006b). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, 27, 107–26.

Clahsen, H. and Felser, C. (2006c). How native-like is non-native language processing? *Trends in Cognitive Sciences*, 10, 564–70.

Clahsen, H., Felser, C., Neubauer, K., Sato, M., and Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning*, 60, 21–43.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181–204.

Clark, H. H. and Wilkes-Gibbs, D. (1989). Referring as a collaborative process. *Cognition*, 22, 1–39.

Clifton, C., Jr., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., and Rayner, K. (2003). The use of thematic role information in parsing: syntactic processing autonomy revisited. *Journal of Memory and Language*, 49, 317–34.

Crain, S. and McKee, C. (1985). Acquisition of structural restrictions on anaphora. *Proceedings: North American Linguistic Society*, 16, 94–110.

Daneman, M. and Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–66.

De Smedt, K. (1990). IPF: an incremental parallel formulator. In R. Dale, C. Mellish, and M. Zock (eds), *Current Research in Natural Language Generation* (pp. 167–92). London: Academic Press.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283.

Eberhard, K. M., Cutting, J. C., and Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological Review*, 112, 531–59.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195–225.

Engelhardt, P. E., Bailey, K. G. D., and Ferreira, F. (2006). Do speakers and hearers observe the Gricean maxim of quantity? *Journal of Memory and Language*, 54, 554–73.

Engle, R. W. (2010). Role of working-memory capacity in cognitive control. *Current Anthropology*, 51, S17–S26.

Epstein, S. D. and Seely, T. D. (2006). *Derivations in Minimalism* (Vol. 111). Cambridge: Cambridge University Press.

Fausey, C. M. and Boroditsky, L. (2010). Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review*, 17, 644–50.

Fausey, C. M. and Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, 18, 150–7.

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology*, 44, 491–505.

Fedorenko, E., Behr, M. K., and Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, 108, 16428–33.

Fedorenko, E. and Thompson-Schill, S. L. (2014). Reworking the language network. *Trends in Cognitive Sciences*, 18, 120–6.

Felser, C. and Roberts, L. (2007). Processing wh-dependencies in a second language: a cross-modal priming study. *Second Language Research*, *23*, 9–36.

Felser, C., Roberts, L., Marinis, T., and Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24, 453–89.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47, 164–203.

Ferreira, F. (2005). Psycholinguistics, formal grammars, and cognitive science. *The Linguistic Review*, 22, 365–80.

Ferreira, F., Bailey, K. G. D., and Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11–15.

Ferreira, F., Christianson, K., and Hollingworth, A. (2001). Misinterpretations of garden-path sentences: implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*, 30, 3–20.

Ferreira, F. and Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348–68.

Ferreira, F. and Patson, N. D. (2007). The "good enough" approach to language comprehension. *Language and Linguistics Compass*, 1, 71– 83.

Ferreira, F. and Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46, 1, 57–84.

Ferreira, V. S. and Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40, 4, 296–340.

Ferreira, V. S. and Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research*, 32, 6, 669–92.

Ferreira, V. S., Bock, K., Wilson, M. P., and Cohen, N. J. (2008). Memory for syntax despite amnesia. *Psychological Science*, 19, 9, 940–6.

Fodor, J. A. (1983). *Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, 9, 427–73.

Fodor, J. D. (1989). Empty categories in sentence processing. *Language and Cognitive Processes*, 4, SI155–SI209.

Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics*, 32, 187–220.

Frazier, L. (1978). *On Comprehending Sentences: Syntactic Parsing Strategies*. PhD dissertation, University of Connecticut, Storrs, CT. Distributed by the Indiana University Linguistics Club, Bloomington, IN.

Frazier, L. (1987). Sentence processing: a tutorial review. In M. Coltheart (ed.), *Attention and Performance 12: The Psychology of Reading* (pp. 559–86). Hillsdale, NJ: Erlbaum.

Frazier, L. (1998). Getting there (slowly). *Journal of Psycholinguistic Research*, 27, 123–46.

Frazier, L. and Flores D'Arcais, G. B. (1989). Filler-driven parsing: a study of gap filling in Dutch. *Journal of Memory and Language*, 28, 331–44.

Frazier, L. and Fodor, J. D. (1978). The sausage machine: a new two-stage parsing model. *Cognition*, 6, 1–34.

Frazier, L. and Rayner, K. (1982). Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178–210.

Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: a review of methodologies and experimental findings. *Second Language Research*, 21, 175–98.

Friederici, A. D. (1998). Diagnosis and reanalysis: two processing steps the brain may differentiate. In F. Ferreira and J. D. Fodor (eds), *Reanalysis in Sentence Processing* (pp.177–200). Dordrecht: Kluwer.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., and Lotocky, B. (1997). The relative contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93.

Garrett, M. F. (1975). The analysis of sentence production. *Psychology of Learning and Motivation*, 9, 133–77.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 1–76.

Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, and W. O'Neil (eds), *Image, Language, Brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gibson, E. (2006). The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, 54, 363–88.

Gibson, E., Bergen, L., and Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110, 8051–6.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1, 3–55.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., and Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, 1, 23–64.

Glenberg, A. M. and Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review*, 9, 558–65.

Grosjean, F. (2008). *Studying Bilinguals*. London: Oxford University Press.

Grosjean, F. (2010). *Bilingual: Life and Reality*. Boston, MA: Harvard University Press.

Grosjean, F. and Li, P. (2012). *The Psycholinguistics of Bilingualism*. New York: Wiley Blackwell.

Hale, J. (2003). *Grammar, Uncertainty and Sentence Processing.* PhD dissertation, Johns Hopkins University, Baltimore, MD.

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., and Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: evidence from written and spoken dialogue. *Journal of Memory and Language*, 58, 2, 214–38.

Hartsuiker, R. J., Pickering, M. J., and Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish–English bilinguals. *Psychological Science*, 15, 6, 409–14.

Havik, E., Roberts, L., Van Hout, R., Schreuder, R., and Haverkort, M. (2009). Processing subject-object ambiguities in the L2: a self-paced reading study with German L2 learners of Dutch. *Language Learning*, 59, 73–112.

Haywood, S. L., Pickering, M. J., and Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science*, 16, 362–6.

Hirsh-Pasek, K. and Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: a window onto emerging language comprehension. In D. McDaniel, C. McKee, and H. S. Cairns (eds), *Methods for Assessing Children's Syntax* (pp. 105–24). Cambridge, MA: MIT Press.

Hoeks, J. C., Hendriks, P., Vonk, W., Brown, C. M., and Hagoort, P. (2006). Processing the noun phrase versus sentence coordination ambiguity: thematic information does not completely eliminate processing difficulty. *The Quarterly Journal of Experimental Psychology*, 59, 1581–99.

Hoff, E. (2012). *Research Methods in Child Language: A Practical Guide.* London: Wiley Blackwell.

Hoff, E. (2014). *Language Development*, 5th edn. Belmont, CA: Wadsworth.

Hohne, E. A. and Jusczyk, P. W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception & Psychophysics*, 56, 613–23.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22, 369–97.

Hussey, E. K. and Novick, J. M. (2012). The benefits of executive control training and the implications for language processing. *Frontiers in Psychology*, 3, 158.

Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research*, 1146, 2–22.

Jackendoff, R. (2002). *Foundations of Language*. New York: Oxford University Press.

Jackendoff, R. (2011). What is the human language faculty?: Two views. *Language*, 87, 586–624.

Jaeger, H. (2007). Echo state network. *Scholarpedia*, 2, 2330.

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–46.

January, D., Trueswell, J. C., and Thompson-Schill, S. L. (2009). Co-localization of stroop and syntactic ambiguity resolution in Broca's area: implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, 21, 2434–44.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25, 603–34.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, 57, 1–33.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., and Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32, 402–20.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, 3, 323–8.

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, 87, 329.

Just, M. A. and Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, 98, 122–49.

Kantola, L. and van Gompel, R. P. G. (2011). Between- and within-language priming is the same: evidence for shared bilingual syntactic representations. *Memory and Cognition*, 39, 276–90.

Kempen, G. and Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11, 2, 201–58.

Kennison, S. M. (2001). Limitations on the use of verb information in sentence comprehension. *Psychonomic Bulletin and Review*, 8, 132–8.

Kim, A. and Osterhout, L. (2005). The independence of combinatory semantic processing: evidence from event-related potentials. *Journal of Memory and Language*, 52, 205–25.

Kim, J.-H. and Christianson, K. (2013). Sentence complexity and working memory effects in ambiguity resolution. *Journal of Psycholinguistic Research*, 42, 393–411.

Kimball, J. (1975). Predictive analysis and over-the-top parsing. In J. Kimball (ed.), *Syntax and Semantics*. New York: Academic Press.

Kolk, H. H. J., Chwilla, D. J., van Herten, M., and Oor, P. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain and Language*, 85, 1–36.

Konopka, A. E. (2012). Planning ahead: how recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, 66, 143–62.

Konopka, A. E. and Brown-Schmidt, S. (2014). Message encoding. In M. Goldrick, V. Ferreira, and M. Miozzo (eds), *The Oxford Handbook of Language Production* (pp. 3–20). Oxford: Oxford University Press.

Kooijman, V., Hagoort, P., and Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, 24, 109–16.

Kraljic, T. and Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: for the speaker or for the addressee? *Cognitive Psychology*, 50, 194–231.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Research*, 1146, 23–49.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., and Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes*, 21, 489–530.

Kutas, M. and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207, 203–5.

Kutas, M. and Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, 11, 539–50.

Kutas, M., Van Petten, C. K., and Kluender, R. (2006). Psycholinguistics electrified II: 1994–2005. In M. Traxler and M. A. Gernsbacher (eds), *Handbook of Psycholinguistics* (pp. 659–724). New York: Elsevier.

Larsen-Freeman, D., Long, M. H., and Jiang, Z. (1991). *An Introduction to Second Language Acquisition Research*. London: Longman.

Laszlo, S. and Federmeier, K. D. (2009). A beautiful day in the neighborhood: an event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, 61, 3, 326–38.

Legate, J. A., Pesetsky, D., and Yang, C. (2014). Recursive misrepresentations: a reply to Levinson (2013). *Language*, 90, 515–28.

Levelt, W. J. M. (1998). *Speaking: From Intention to Articulation*. Boston, MA: MIT Press.

Levelt, W. J. M. (2012). *A history of Psycholinguistics: The Pre-Chomskyan Era*. Oxford: Oxford University Press.

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1, 1–38.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126–77.

Levy, R., Bicknell, K., Slattery, T., and Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, 106, 21086–90.

Lewis, R. and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375–419.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.

Lim, J.-H. and Christianson, K. (2013a). Integrating meaning and structure in L1–L2 and L2–L1 translations. *Second Language Acquisition*, 29, 233–56.

Lim, J.-H. and Christianson, K. (2013b). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition*, 16, 518–37.

Lim, J.-H. and Christianson, K. (2014). L2 Sensitivity to agreement errors: evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, 36, 6, 1–33.

Liversedge, S. P., Pickering, M. J., Branigan, H. P., and van Gompel, R. P. (1998). Processing arguments and adjuncts in isolation and context: the case of *by*-phrase ambiguities in passives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 2, 461.

Lotto, A. J., Hickok, G. S., and Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 13, 110–14.

Luke, S. G. and Christianson, K. (2012). Semantic predictability eliminates the transposed-letter effect: insights from a combined self-paced reading and masked-priming paradigm. *Memory & Cognition*, 40, 628–41.

MacDonald, M. C., Just, M. A., and Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24, 56–98.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.

MacKay, D. G. (1972). The structure of words and syllables: evidence from errors in speech. *Cognitive Psychology*, 3, 2, 210–27.

Marinis, T., Roberts, L., Felser, C., and Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, 27, 53–78.

Marslen-Wilson, W. and Tyler, L. K. (1975). Processing structure of sentence perception. *Nature*, 257, 784–6.

McDaniel, D., McKee, C., and Cairns, H. S. (1996). *Methods for Assessing Children's Syntax*. Cambridge: Cambridge University Press.

McDonald, J. L. (2006). Beyond the critical period: processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55, 381–401.

McMurray, B., Aslin, R. N., and Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, 12, 369–78.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, 7, 141–4.

Miller, G. A. and Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 2, 217–28.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17, 2, 357–74.

Nicol, J. and Pickering, M. J. (1993). Processing syntactically ambiguous sentences: evidence from semantic priming. *Journal of Psycholinguistic Research*, 22, 207–237.

Nicol., J. and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18, 5–19.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., and Bunting, M. F. (2014). Clearing the garden-path: improving sentence processing through cognitive control training. *Language, Cognition, and Neuroscience*, 29, 186–217.

Novick, J. M., Trueswell, J. C., and Thompson-Schill, S. L. (2005). Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, and Behavioral Neuroscience*, 5, 263–81.

Onifer, W. and Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: effects of frequency of meaning and contextual bias. *Memory & Cognition*, 9, 225–36.

Papadopoulou, D. and Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing. *Studies in Second Language Acquisition*, 25, 501–28.

Patson, N. D., Darowski, E. S., Moon, N., and Ferreira, F. (2009). Lingering misinterpretations in

garden-path sentences: evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 280–5.

Payne, B. R., Grison, S., Gao, X., Christianson, K., Morrow, D. G., and Stine-Morrow, E. A. (2014). Aging and individual differences in binding during sentence understanding: evidence from temporary and global syntactic attachment ambiguities. *Cognition*, 130, 157–73.

Perea, M. and Lupker, S. J. (2003). Does JUGDE activate COURT? Transposed-letter confusability effects in masked associative priming. *Memory & Cognition*, 31, 829–41.

Perea, M. and Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, 51, 231–46.

Phillips, C. (2006). The real-time status of island phenomena. *Language*, 82, 795–82.

Phillips, C., Wagers, M. W., and Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (ed.), *Experiments at the Interfaces, Syntax & Semantics*, vol. 37 (pp. 153–186). Bingley: Emerald Publications.

Pickering, M. J. and Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological Bulletin*, 134, 3, 427.

Pickering, M. J. and Traxler, M. J. (1998). Plausibility and recovery from garden paths: an eye- tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 940–61.

Pickering, M. J., Traxler, M. J., and Crocker, M. W. (2000). Ambiguity resolution in sentence processing: evidence against frequency-based accounts. *Journal of Memory and Language*, 43, 447–75.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, 77, 2, 97–132.

Pulman, S. G. (1986). Grammars, parsers and memory limitations. *Language and Cognitive Processes*, 1, 197–225.

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Science*, 17, 458–70.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372.

Rayner, K. (2009). The thirty fifth Sir Frederick Bartlett Lecture: eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62, 1457–506.

Rayner, K., Carlson, M., and Frazier, L. (1983). The intersection of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358–74.

Rayner, K., Pollatsek, A., Ashby, J., and Clifton, C. (2012). *The Psychology of Reading*, 2nd edn. New York: Psychology Press.

Rayner, K. and Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: a further examination. *Psychonomic Bulletin & Review*, 3, 504–9.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 275.

Reichle, E. D., Rayner, K., and Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: comparisons to other models. *Behavioral and Brain Sciences*, 26, 445–76.

Rogalsky, C. and Hickok, G. (2011). The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience*, 23, 1664–80.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–8.

Schlesewsky, M. and Bornkessel, I. (2006). Context-sensitive neural responses to conflict resolution: electrophysiological evidence from subject–object ambiguities in language comprehension. *Brain Research*, 1098, 139–52.

Scott, R. M. and Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, 122, 163–80.

Seidl, A., Cristià, A., Bernard, A., and Onishi, K. H. (2009). Allophonic and phonemic contrasts in infants' learning of sound patterns. *Language Learning and Development*, 5, 191–202.

Seidl, A. and Cristià, A. (2012). Infants' learning of phonological status. *Frontiers in Psychology*, 3, 448.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., and Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words: evidence from priming and eye fixations. *Memory & Cognition*, 28, 1098–108.

Senay, I., Albarracín, D., and Noguchi, K. (2010). Motivating goal-directed behavior through introspective self-talk: the role of the interrogative form of simple future tense. *Psychological Science*, 21, 499–504.

Shin, J.-A. and Christianson, K. (2009). Syntactic processing in Korean–English bilinguals: evidence from cross-linguistic structural priming. *Cognition*, 112, 175–80.

Shin, J.-A. and Christianson, K. (2011). The status of dative constructions in Korean, English, and in the minds of Korean–English bilinguals. In H. Yamashita, J. Packard, and Y. Hirose (eds), *Processing and Producing Head-Final Structures* (pp. 153–69). New York: Springer.

Simon, H. A. (1947). *Administrative Behavior*. New York: Macmillan.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., and Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from flawed semantic processing. *Journal of Memory & Language*, 69, 104–20.

Solomon, E. S. and Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology*, 49, 1–46.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., and Sedivy, J. C. (2002). Eye movements and spoken language comprehension: effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45, 447–81.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 116, 71–86.

Staub, A. and Clifton, C., Jr. (2006). Syntactic prediction in language comprehension: evidence from *either… or*. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 425–36.

Staub, A. and Rayner, K. (2007). Eye movements and on-line comprehension processes. In M. G. Gaskell (ed.), *The Oxford Handbook of Psycholinguistics* (pp. 327–42). Oxford: Oxford University Press.

Swets, B., Desmet, T., Hambrick, D. Z., and Ferreira, F. (2007). The role of working memory insyntactic ambiguity resolution: a psychometric approach. *Journal of Experimental Psychology*, 136, 64–81.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 3617–32.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645–59.

Tanenhaus, M. K. (2007). Eye movements and spoken language processing. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, and R. L. Hill (eds), *Eye Movements: A Window on Mind and Brain* (pp. 309–26). Oxford: Elsevier.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–4.

Tanner, D. (2013). Individual differences and streams of processing. *Linguistic Approaches to Bilingualism*, 3, 350–6.

Tokowicz, N. and Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: a self-paced reading study. *European Journal of Cognitive Psychology*, 22, 1092–106.

Travis, K. E., Leonard, M. K., Brown, T. T., Hagler, D. J., Curran, M., … Halgren, E. (2011). Spatiotemporal neural dynamics of word understanding in 12- to 18-month-old-infants. *Cerebral Cortex*, 21, 1832–9.

Traxler, M. J. (2009). A hierarchical linear modeling analysis of working memory and implicit prosody in the resolution of adjunct attachment ambiguity. *Journal of Psycholinguistic Research*, 38, 491–509.

Traxler, M. J. (2012). *Introduction to Psycholinguistics: Understanding Language Science*. Malden, MA: Wiley Blackwell.

Traxler, M. J., Williams, R. S., Blozis, S. A., and Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, 53, 204–24.

Trott, K., Dobbinson, S., and Griffiths, P. (2004). *The Child Language Reader*. New York: Psychology Press.

Trueswell, J. C. and Gleitman, L. R. (2007). Learning to parse and its implications for language acquisition. In G. Gaskell (ed.), *Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press.

Trueswell, J. C., Sekerina, I., Hill, N. M., and Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition*, 73, 89–134.

Trueswell, J. C. and Tanenhaus, M. K. (2005). *Approaches to Studying World-Situated Language Use: Bridging the Language as Product and Language as Action Traditions*. Cambridge, MA: MIT Press.

Trueswell, J. C., Tanenhaus, M. K., and Garnsey, S. M. (1994). Semantic influences on parsing: use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 285–318.

Trueswell, J. C., Tanenhaus, M. K., and Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528–53.

Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 231–70.

Valentine, J. R. (2001). *Nishnaabemwin Reference Grammar*. Toronto: University of Toronto Press.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 407–30.

Van Dyke, J. A. and Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: a cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49, 285–316.

van Gompel, R. P. G. and Pickering, M. J. (2007). Syntactic parsing. In R. Gomez (ed.), *The Oxford Handbook of Psycholinguistics* (pp. 289–307). Oxford: Oxford University Press.

van Gompel, R. P. G., Pickering, M. J., Pearson, J., and Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: evidence from structural priming. *Journal of Memory and Language*, 55, 335–62.

van Herten, M., Kolk, H. H. J., and Chwilla, D. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22, 241–55.

van Nice, K. Y. and Dietrich, R. (2003). Task sensitivity of animacy effects: evidence from German picture descriptions. *Linguistics*, 41, 825–50.

Van Petten, C. and Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83, 176–90.

Vasishth, S. and Drenhaus, H. (2011). Locality in German. *Dialogue & Discourse*, 1, 59–82.

Vasishth, S. and Lewis, R. L. (2006). Argument-head distance and processing complexity: explaining both locality and antilocality effects. *Language*, 82, 767–94.

Warren, T. and Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, 85, 79–112.

Werker, J. F., Yeung, H. H., and Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, 21, 221–6.

White, L. (2003). *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Wilkes-Gibbs, D. and Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31, 183–94.

Williams, J. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, 9, 71–81.

Yamashita, H. and Chang, F. (2001). "Long before short" preference in the production of a head-final language. *Cognition*, 81, B45–B55.

Yang, C. D. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Sciences*, 8, 451–6.

Yuan, S., Fisher, C., and Snedeker, J. (2012). Counting the nouns: simple structural cues to verb meaning. *Child Development*, 83, 1382–99.

Zwaan, R. A. (2004). The immersed experiencer: toward an embodied theory of language comprehension. *Psychology of Learning and Motivation*, 44, 35–62.

Zwaan, R. A. and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162.

18 Sign Languages*

WENDY SANDLER AND DIANE LILLO-MARTIN

It has been over 50 years since serious investigation of sign languages began to show that these languages are bona fide linguistic systems, with structures and rules and the full range of expressive power that characterize spoken languages. Researchers have spent most of that time demonstrating, with increasing rigor and formality, the sometimes surprising similarities between languages in the two modalities, spoken and signed. Concomitantly, scholars in related disciplines like language acquisition and neurolinguistics have been discovering significant similarities between spoken and signed languages in these domains as well. It is safe to say that the academic world is now convinced that sign languages are real languages in every sense of the term.

If this were the whole story, however, there would be no need for a chapter on sign languages in a volume such as this one. Each sign language would be seen as a language like any other, English, Hungarian, Central Alaskan Yupik Eskimo, or Mandarin Chinese, each with its own contribution to make toward understanding the general language faculty of humans. But this is not the whole story. Rather, sign languages as a group are of special importance to our understanding of the essential nature of language, for two reasons. First, the study of natural languages in a different physical modality confirms in a novel way the hypothesis that all natural human languages are characterized by certain key properties. And second, this study raises fundamental questions about the human language capacity, as well as challenges for language theory, that we would never have noticed were it not for the existence of sign languages.

The sign languages under discussion are the languages used by communities of deaf people all over the world. They are natural languages, in the sense that they are not consciously invented by anyone, but rather develop spontaneously wherever deaf people have an opportunity to congregate and communicate regularly with each other. Sign languages are not derived from spoken languages; they have their own independent vocabularies and their own grammatical structures. Although there do exist contrived sign systems that are based on spoken languages (such as Signed English, Signed Hebrew, etc.), such systems are not natural languages, and they

* Sandler's research was supported in part by Israel Science Foundation Grant No. 820/95, Binational Science Foundation Grant No. 9500310/2, National Institute on Deafness and Communication Disorders, National Institutes of Health, Grant No. DC6473, and European Research Council grant No. 340140; and Lillo-Martin's by National Institute on Deafness and Communication Disorders, National Institutes of Health, Grant Nos. R01 DC 00183, and R01 DC 009263, and the University of Connecticut Research Foundation. We would like to thank Mark Aronoff, Peter MacNeilage, Richard Meier, Irit Meir, and Yudit Rosenberg for their helpful comments on earlier versions of this chapter.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

are not the object of interest here. Rather, linguists and cognitive psychologists are most interested in the natural sign languages passed down without instruction from one deaf generation to the next, and used by deaf people in their own communities all over the world.

Sign languages exhibit the full range of expression that spoken languages afford their users. Different styles are adopted for different social contexts; university lectures are delivered in or translated to sign language; storytelling has been heightened to an art in some deaf communities; deaf poets create artistic poetry in signs, marshaling the formational elements of the languages to convey images, emotions, and ideas. Sign languages can do everything that spoken languages can. In this chapter, we briefly explain how they do so.

Most of the examples we provide in our survey of the topics we have introduced here come from American Sign Language (ASL) and Israeli Sign Language (ISL), but the properties they exemplify are common to many sign languages that have been studied so far. This does **not** mean that there is one universal sign language or that all sign languages have the same grammatical structure, though. The common elements described here are selected because they provide useful examples of the linguistic patterning of sign languages in general, and of the ways in which they compare to spoken language patterns.

It is not only through linguistic analysis of adult language competence that we investigate theoretical questions about the nature of human language. Other methods help to complete the picture. The creative use of language in poetry allows us to identify the building blocks of language by isolating and using them in new ways, while at the same time offering insight into deaf culture. Tracing the course of acquisition in children and investigating the neural control of languages in the visual modality are productive research foci. One area of investigation that has attracted a good deal of attention in recent years is the study of young sign languages, many of which have emerged in small, isolated communities with disproportionately high numbers of deaf people. We survey each of these areas briefly here.[1]

1 Linguistic Structure of Sign Languages

First we examine the structure of the sentence (syntax), and then move to the structure of the smaller units of language, those that may be compared to the meaningless but identifiable sounds of speech (phonology). We will end the linguistic description with a discussion of the structure of words (morphology).

1.1 Sentence structure: Syntax

One of the fundamental properties of human language is that it can be used to create an unlimited number of utterances given a limited number of pieces. At the syntactic level, this property follows directly from a mathematical property of language called recursion. We're all familiar with recursion (even if not with the term). It is found in language, and computer programs, and even in children's stories, as in (1).

(1) This is the farmer sowing the corn,
That kept the cock that crowed in the morn,
That waked the priest all shaven and shorn,
That married the man all tattered and torn,
That kissed the maiden all forlorn,
That milked the cow with the crumpled horn,
That tossed the dog,
That worried the cat,

That killed the rat,
That ate the malt
That lay in the house that Jack built.

In (1), the process or rule that creates a relative clause (here, the clauses beginning with *that*) is applied repeatedly to the noun phrases inside other relative clauses. This repeated application of the same rule to create more and more complex sentences is an example of recursion. The children's story is amusing precisely because we all know that there is no theoretical limit to the application of this rule. Any speaker of English can add to the story by generating another relative clause at the beginning, as in (2).

(2) This is the banker, his honor forsworn
That foreclosed on the farmer sowing the corn…,

The only limitations on the number of relative clauses are practical and not linguistic: the speaker may run out of things to say, or out of breath, or time, or memory. It is because the rules of syntax are allowed to apply recursively that language is nonfinite: there is no limit to its expressive power. It is important that this recursion applies to create structures which are **embedded** inside other structures, to create subordinate clauses, such as the "*that*" relative clauses in (2), among other types of recursion. These embedded subordinate clauses involve more complex structure than coordinate clauses, which are illustrated in (3).

(3) This is the dog, and the dog worried the cat, and the cat killed the rat, and the rat ate the malt, and the malt lay in the house and Jack built the house.

Embedding distinguishes subordinate clauses of the sort exemplified by *The House that Jack Built* from simple coordination. A simplified diagram of the structure of recursive subordinate versus coordinate clauses is given in Figure 18.1. The rule that makes a relative clause – put a sentence inside a noun phrase – produces a structure in which one sentence is inside another sentence. This is the key to recursion: by putting one phrase inside another of the same type, there is in principle no limit to the length of a sentence.

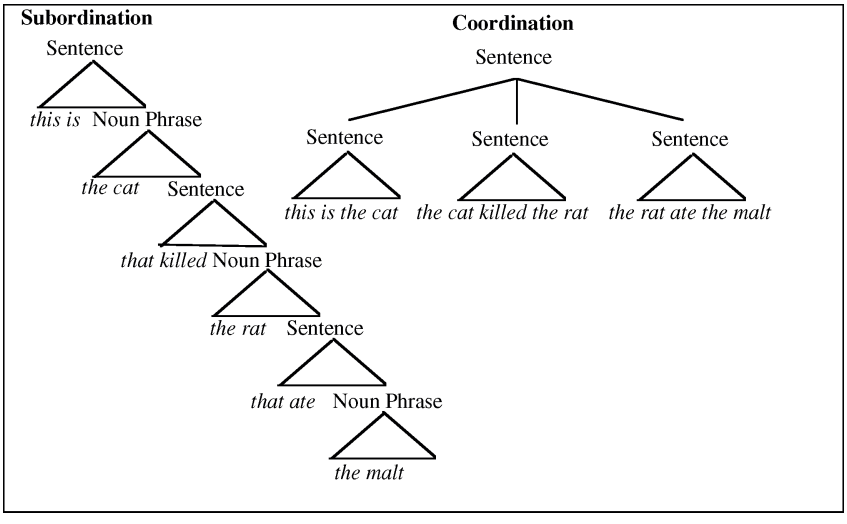


Figure 18.1 Simplified phrase structure diagrams of subordination and coordination.
© Wendy Sandler and Diane Lillo-Martin.

Human language is not limited to simple conjunction of one phrase after another (like that in example (3)); it has both coordination and subordination.[2] Thus, it might be rather surprising that it was claimed in the early days of the study of ASL that it actually lacks subordination. Thompson (1977) attempted to discover the mechanisms for subordination in ASL, and, not finding what he considered to be evidence for it, decided that it was systematically missing. If this were correct, it might be concluded that sign languages are incomplete or less rich in structure compared to spoken languages.

Later research has shown that it would be inappropriate to draw such conclusions. Liddell (1980) argued convincingly that Thompson's claim was incorrect. He showed that ASL has subordination, even if the markers of this process are different in ASL from those in English (see Section 18.1.2). Liddell's analysis has been substantiated by many researchers since.

Padden (1988) confirmed Liddell's argument for subordination in ASL. She provided a number of structural tests which distinguished between embedded and coordinate clauses. For example, she looked closely at a process by which a pronoun copy of the subject of a clause can be tagged onto the end of the sentence, a process she called Subject Pronoun Copy. As example (4) (modified from Padden 1988) shows, a sentence-final pronoun may be a copy of the first subject in a complex utterance only if the utterance contains an embedded clause – not a coordinate clause. (In this example and others, individual signs are represented using uppercase English glosses that are rough translation equivalents. INDEX refers to the pointing sign used as a pronoun, with a subscript indicating its reference.) In (4a), which contains an embedded clause, the sentence-final subject pronoun copy ($_{i}$INDEX) can refer back to the main or matrix clause subject, MOTHER. Example (4b), on the other hand, contains a coordinate clause. The subject pronoun copy ($_{1}$INDEX), referring back to the subject of the first clause ('I'), is not allowed. Example (4c) shows that a pronoun copy is allowed in a coordinate structure, but the pronoun must refer to the subject of the second clause (BROTHER). In other words, the process of pronoun copy applies differently to coordinate and embedded clauses, allowing a sentence-final INDEX to corefer with the initial subject only in the latter case. This contrast reveals that there are embedded sentences in ASL, a form of recursion.

(4) a. $_{i}$MOTHER SINCE $_{i}$PERSUADE$_{j}$ $_{j}$BROTHER $_{j}$COME$_{k}$ $_{i}$INDEX.
'My mother has been urging my brother to come and stay here, she has.'

b. * $_{1}$INDEX $_{1}$HIT$_{i}$ $_{i}$BROTHER, $_{i}$INDEX TATTLE $_{j}$MOTHER $_{1}$INDEX.
'I hit brother and he told mother, I did.'

c. $_{1}$INDEX $_{1}$HIT$_{i}$ $_{i}$BROTHER, $_{i}$INDEX TATTLE $_{j}$MOTHER $_{i}$INDEX.
'I hit brother and he told mother, he did.'

A controversy over word order properties in ASL was similarly resolved by more careful analysis, and by looking beyond English. Since basic word order is very flexible in ASL, some early researchers argued that only pragmatic considerations, such as emphasis of some part of the sentence, and not syntactic rules dictate word order in ASL sentences. These researchers proposed that there are no syntactic rules for ordering words in ASL. However, since then, many researchers have collected evidence to support the point of view that ASL – like Japanese, Russian, and other spoken languages which allow for relative flexibility in surface word order – has an underlying structure and word order which can be modified by rules of the grammar. In the case of ASL, the underlying word order is Subject-Verb-Object. Here, too, most researchers now assume the basic order argued for in these works, and current work concentrates on the rules and principles which generate this order and its many variants.

Using the tools of linguistic analysis, it is possible to go far beyond the observation that ASL has recursion and subordination and a basic word order to look for other universal characteristics.

An influential hypothesis in linguistics is the Universal Grammar Hypothesis, according to which there are constraints that are operative across languages, even if overt evidence in the linguistic environment does not seem to be available for them. If some knowledge which adults are shown to possess is not accessible to the language learner, it is hypothesized to be part of Universal Grammar. Consider the sentences in (7).

(7) a. Steve likes beer with pizza.
b. What does Steve like beer with?
c. Yoav likes wine and cheese.
d. *What does Yoav like wine and?

Although the relationship between (7a) and (7b) is very similar to the relationship between (7c) and (7d), only the first pair is grammatical. Making a question out of the noun phrase following a preposition is (usually) grammatical in (colloquial) English, but making a question out of a noun phrase which is coordinated with another noun phrase is not. In fact, the restriction on questions like (7d), which disallows extraction of material out of a coordinated structure, is virtually universal, and it has been proposed that a general universal constraint – the Coordinate Structure Constraint (Ross 1967) – prohibits it.

Whether or not one accepts the hypothesis that such constraints are innately specified, the fact that they are not easily deducible from the input but appear to be ubiquitous in spoken languages makes them key exemplars of the human language capacity. It is therefore important to determine whether or not they hold for sign languages as well. If so, this would be important evidence that the constraints are truly universal, and that sign languages have the same properties as any other natural language. In fact, several researchers have argued that this is the case.

The Coordinate Structure Constraint demonstrated in (7) provides the clearest example. Padden (1988) shows that coordinate structures are allowed in ASL, as illustrated in (8) and (9).

(8) $_{i}$INDEX $_{i}$GIVE$_{1}$ MONEY, $_{j}$INDEX $_{j}$GIVE$_{1}$ FLOWERS
'He gave me money but she gave me flowers.'

_____t _____t

(9) *FLOWER, $_{i}$GIVE$_{1}$; MONEY, $_{j}$GIVE$_{1}$
'Flowers, he gave me money but she gave me.'

ASL, like English, clearly prohibits violations of the Coordinate Structure Constraint, as illustrated in (9). The 't' on the line over FLOWER in (9) indicates nonmanual marker for topicalization, the process by which the topic, FLOWER, is extracted and moved out of its original position in the coordinated structure, to the beginning of the sentence, similar to 'what' in (7d). This movement out of a coordinate structure is ungrammatical in ASL as it is in English and other languages.

In several domains of syntax, the constraints proposed to be universal (including the Coordinate Structure Constraint) can be demonstrated to apply to ASL as well as to spoken languages, and it is likely that other signed languages will also show adherence to these constraints.

1.2 The structure of sounds and their sign language equivalents: Phonology

In order to have sentences, one must have words, and words – at least in spoken language – are pronounced as a series of sounds. What about the signs of sign language? Do they have a level of substructure like the spoken word? Since spoken and signed languages are produced and

perceived by different physical systems (oral/aural, and manual/visual) one might expect to find the least amount of similarity across the two modalities at this level of analysis. Yet, here, too, there is much common ground.

In 1960, William Stokoe published a monograph in which he demonstrated that the words of American Sign Language are not holistic gestures, but rather are analyzable as a combination of three meaningless yet linguistically significant categories: handshapes, locations, and movements. That is, by changing some feature of any one of those three categories in a sign language, themselves meaningless, one could change the meaning of the sign. For example, in Israeli Sign Language (ISL), by changing only the configuration of the hand, the signs MOTHER and NOON are distinguished. In these two signs, the locations and movements are the same. Only the hand configuration is different. Similar pairs exist that are distinguished only by their locations or only by their movements.

The example in Figure 18.2 is analogous to the English pair, *pan, tan*, in which the first sound of each word – *p* and *t* – is different. The sounds are themselves meaningless, but they are linguistically significant because they make a difference in meaning when put in a word. In the sign language pair, MOTHER, NOON, the hand configurations are also meaningless, yet they too make a difference in meaning when combined with other parts of a sign. The other formational elements – locations and movements – can, like hand configurations, independently make a difference in meaning, though they pattern in sign language phonology according to form, rather than meaning.

This finding was of supreme importance. Ever since its discovery, it has no longer been possible to assume, as most people previously had, that signs are fundamentally different from spoken words, that they are simple iconic gestures with no substructure. Rather, Stokoe showed that ASL is characterized by a fundamental feature of language in general: a property that Charles Hockett (1960) called duality of patterning. This duality is between the meaningful level (consisting of morphemes, words, phrases, sentences) and the meaningless level, which in spoken languages is the level of the sounds that make up the meaningful expressions. The meaningless elements of spoken language are linguistically significant (i.e., they independently make a difference in meaning); they obey constraints on their combination within morphemes and words; and they may be systematically altered in different phonetic contexts. This is the domain of phonology. The list of handshapes, locations, and movements are the

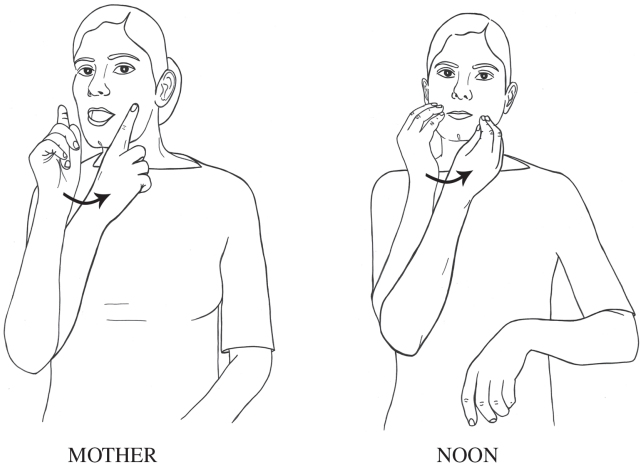


Figure 18.2 ISL minimal pair. © Wendy Sandler and Diane Lillo-Martin.

formational elements of sign language phonology, comparable to the list of phones in spoken language. We will now show that sign language phonology is also characterized by constraints on the combination of these elements, and by systematic changes in "pronunciation" under certain circumstances.

All languages have constraints on the co-occurrence of sounds in syllables and words. For example, English does not allow the sequences **sr* or **chl* at the beginning of a syllable or word (although other languages do permit such combinations). Sign languages as well have constraints on the combination of elements at this same level of structure. For example, only one group of fingers may characterize the handshape within any morpheme (Mandel 1981). While either the finger group 🖐 (all fingers) or the group ✌ (index plus middle finger) may occur in a morpheme, a sequence of the two shapes, * 🖐✌ is prohibited within a morpheme in the native signs of ASL and other sign languages.

Similarly, all languages have assimilation processes, in which sounds borrow some or all aspects of neighboring sounds. For example, in the English compound words, *greenback* and *beanbag*, the nasal sound [n] often assimilates the labial place of articulation from the [b] that follows it: *gree[m]back, bea[m]bag*. In many common ASL compounds, part of the hand configuration – its orientation – may similarly assimilate from one part of the compound to the other. The example in Figure 18.3 is from the compound which means BELIEVE, made from the two words THINK and MARRY. Just as the [n] borrowed one of the features of [b] (the bilabial feature) in the English example above, in the ASL compound, the hand configuration of THINK borrows a feature – the orientation feature – from the following sign in the compound, MARRY. That is, rather than being oriented toward the face as in the citation form of THINK, the dominant, signing hand in the compound BELIEVE is oriented outwards, as in the sign, MARRY, in both parts of the compound (Sandler 1989).

While there are many differences in the phonologies of signed and spoken languages (Brentari 1998; Sandler in press), the discovery of similarities in such a different physical channel are impressive, some of them at even more surprising levels of analysis than minimal pairs. For example, it has been demonstrated that the phonological elements of ASL words are not all simultaneously organized as Stokoe had claimed, but rather have a certain amount of sequential structure, as spoken languages have one sound after another. A sign language equivalent of the syllable has even been argued for.

An aspect of language structure that involves both phonology and syntax is prosody. Prosody involves rhythm, to separate the parts of a sentence; prominence, to emphasize selected elements; and intonation, to communicate other important information, such as the discourse function of the sentence, e.g., whether an utterance is a plain declarative sentence or a question. A number of investigators working on different sign languages have shown that sign languages have the equivalent of prosody (e.g., Wilbur 2000; Sandler 2011, 2012b). While spoken languages use the rise and fall of the pitch of the voice, volume, and pause to achieve these effects, sign languages employ facial expressions, body postures, and manual rhythmic patterns in similar ways and for similar functions. Examples of sign language "intonation" are the Israeli Sign Language facial expressions for yes/no questions, and for "shared information," shown in Figure 18.4. These grammatical facial expressions, the equivalent of intonation in sign language, are temporally aligned with rhythmic features like lengthening and pause that separate one constituent from another (Nespor and Sandler 1999). Together, rhythm and intonation help to determine where the boundary is between constituents, which can be a very useful tool in sign language analysis.

Sign language linguistic "intonation" is **different** from the facial expressions used by hearing people in their communication, which are affective and not mandatory or systematic. Rather, linguistic facial expressions of sign language are like the intonational pitch patterns of spoken language. Both tonal melodies and facial melodies are **grammaticalized**: they are conventionalized and systematic. For example, the intonational melody used in spoken language to ask a

Figure 18.3 ASL signs THINK and MARRY, and the compound BELIEVE with orientation assimilation. © Wendy Sandler and Diane Lillo-Martin.

question requiring an answer of "yes" or "no" is systematically different from the one used to make a declarative statement. The same is true of the facial intonations for these two types of sentences in sign language.

The phonology and prosody of signed and spoken languages are clearly not identical to one another. Perhaps the most salient difference is the degree to which sign language accommodates simultaneity of structure: handshape and place of articulation are simultaneously produced in a sign; different facial articulators and manual articulators simultaneously encode prosodic structure. The articulatory features that make up the systems in the two modalities are obviously not

Figure 18.4 Facial expressions accompanying an ISL Y/N question (raised brows, head forward) and "shared information" (squint). © Wendy Sandler and Diane Lillo-Martin.

the same: the feature [voice], universal in spoken languages, cannot characterize a sign, and the feature [index finger] will never be included in the phonological representation of a spoken word. Modality differences such as these should not be ignored (see Sandler in press), as they inform us about the relationship between the production system and language form in both modalities. Yet considering the profound differences between the two physical channels, we are still impressed by similarities in organization at this level (Brentari 2011; Sandler 2012a; Berent 2013).

In the next subsection, the element that is perhaps the most central aspect of language is examined: the word.

1.3 Word structure: Morphology

Most languages have both simple words, such as *teach*, and complex words, such as *teach+er*. Knowing a language entails understanding the internal structure of its complex words, as well as the ability to create and understand new complex words that exploit those same kinds of internal structures. The study of the internal structure of words is called morphology.

For example, given a new English verb *scaff*, as in *The purpose of this machine is to scaff computers against viruses*, we can also create or analyze the internal structure of the word *scaffer* and can deduce something about its meaning in the sentence, *The company purchased several expensive scaffers last year.* We would also immediately judge the nonce word **er+scaff* to be impossible in English. Speakers of English know the form and function of the meaningful word component *-er*, and they know that it is a suffix rather than a prefix. Although users of a language are usually not conscious of their linguistic knowledge, their use of language clearly reveals the existence of this underlying system.

Another type of complex word formation can be thought of as relating words to each other within a sentence. For example, the word *walk* has two different suffixes in the following sentences, *-s*, and *-ed*. The *-s* in sentence (10a) is an agreement marker; it shows that the subject of the sentence, *Hadar*, is third person singular ("he" or "she"). The *-ed* in sentence (10b) is a tense marker, showing that the event described by the sentence took place in the past.

(10) a. Hadar walks to school sometimes.
b. Stephanie walked to her friend's house.

The existence of complexity of structure within words is a typical property of spoken language, though many different kinds of word formation can be found in the languages of the world. Most (though not all) languages have complex words, and many languages have far more complex morphology than English has.

What about sign languages? While one might expect any versatile communication system to have syntax, one might not necessarily expect sign languages to have internal structure to their words. Rather, one might expect, as naive early descriptions of sign language used to claim, that signs are holistic gestures, each one representing a unitary concept. Concomitant with this view is the belief that the vocabulary of sign languages is purely iconic, that there is a one-to-one relationship between the form of a word or concept and its meaning. The ASL sign for AIRPLANE looks something like an airplane; the sign for GIVE looks something like the act of handing something to someone. If these sign words are iconic wholes, then adding grammatical complexity to them in a systematic way might seem counterintuitive. We have already shown that sign words have a meaningless level of internal structure, that they have phonology. We now show that they also have meaningful units that go together to make complex words.

Sign languages have a great deal of morphological complexity. Such complexity is one of many sources of evidence that systematic grammatical structuring strongly influences the presumably iconic origins of these languages. We will describe two kinds of word complexity here: verb agreement and verbs of motion and location.

We begin with verb agreement (Fischer and Gough 1978; Padden 1988). In many spoken languages, verbs have some kind of marking on them that gives information about their subjects, objects, or other nouns directly related to them in the sentence. Example (10a) above shows the only agreement marker that English has on main verbs, the *-s* which indicates third person and singular in the present tense. Other languages have far richer agreement systems. For example, Swahili has both subject and object agreement markers that indicate person ('I', 'you', 's/he'), gender ('he' or 'she'), and number (singular or plural).

(11) Swahili

a. à - ta - ku - penda
he (subj) future you (obj) like
atakupenda – 'he will like you'

b. u - ta - m - penda
you (subj) future him (obj) like
utampenda 'you will like him'

All established sign languages investigated so far show a comparable kind of verb agreement. Consider, for example, the Israeli Sign Language verb SHOW, shown in Figure 18.5.[3] To say 'I show you', the motion of the sign is from a point near the signer toward the addressee. To say 'you show me', the beginning and endpoints of the sign are just the opposite, beginning at a point near the addressee, and ending near the signer. The beginning and endpoints of the sign are markers for the subject and object of the verb it represents. To say, 'I show you (plural)', the hands moves in a horizontal arc in front of the signer.

In the first example, 'I show you', the first position of the hand corresponds to the prefix *a* in the Swahili example in (11) above: it marks agreement with the person of the verb's subject – third person ('he') in the Swahili example, and first person ('I') in the ISL example. The second position of the hand corresponds to the morpheme *ku* in the same Swahili example, agreeing with the person of the object of the verb – second person ('you') in Swahili, and second person also in ISL. The direction of movement, determined by the beginning and endpoints of the second example in Figure 18.5, similarly marks agreement with subject and object – here, 'you' and 'me'. To agree with the second person plural – 'I show you (plural)', the shape of the movement is altered.

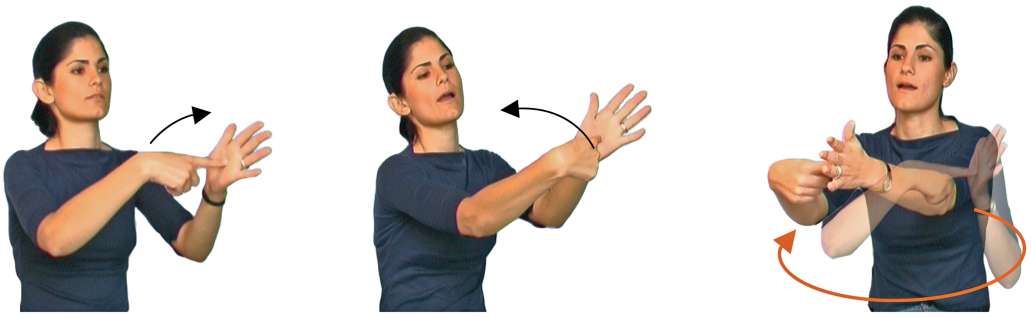


Figure 18.5 ISL verb agreement. © Wendy Sandler and Diane Lillo-Martin.

This kind of phenomenon can be described as subject-object agreement; in particular, sign language verbs agree for person and number of their subject and object. In this way, the verb agreement found in sign languages is similar to that in many spoken languages. (See Lillo-Martin and Meier 2011, along with commentaries and responses, for discussion of the similarities and differences between what we call here sign language agreement and the forms of agreement found in spoken languages.)

A characteristic of verb agreement systems in sign languages is that different categories of verbs participate in this system in different ways. For example, in addition to the subject-object agreement described earlier, some verbs, commonly called backwards verbs, have the opposite agreement pattern of the one shown above. In these verbs, the movement of the hand is from the object to the subject, instead of the usual direction from subject to object. This class includes verbs such as INVITE, TAKE, COPY, ADOPT. The list of verbs showing backwards agreement overlaps to a high degree in ASL and ISL, and possibly across all established sign languages (Meir 2002). Other verbs agree with points in space denoting specific locations, rather than with the verb's subject and object. Many others do not agree at all.

A more complex type of morphology in sign languages is found in a type of construction called classifier predicates. In these constructions, handshapes that stand for classes of nouns combine with different types and manners of movements and with locations in space. The handshapes of classifier constructions often classify nominal referents according to physical properties like size or shape. These complex forms differ from the morphologically simple signs of the language exemplified in the phonology subsection above, because, in the classifier subsystem, each of the formational parameters – handshape, location, and movement – is meaningful, and each hand can represent a separate morpheme as well. In fact, classifier predicates in sign languages, which we will exemplify below, are so visually and spatially motivated that one might be tempted to think that they are different from anything found in spoken language. Yet, despite certain nontrivial differences, even these constructions can be fruitfully compared with structures found in many spoken languages. For example, Digueño, a Hokan-Siouan language spoken in California and Mexico, signifies physical properties of objects with classifier affixes, as shown in (12).

Digueño (Langdon 1970: 78)

(12)

a*′mi…*	'to hang (a long object)'
tu*mi…*	'to hang (a small round object)'
a*·xi¬…*	'to drag (a long object)'
c*′ xi¬…*	'to drag (a bunch of objects)'

In this language, the prefix, printed in bold in (12), is a classifier of some physical property of a nominal: a LONG OBJECT, a SMALL ROUND OBJECT, or a BUNCH OF OBJECTS, in these examples.

Studies that have compared sign language and spoken language classifiers have found parallels in the categories that the forms classify in the two language modalities (Supalla 1986). Size and Shape Classifiers, or SASSes, one of three major kinds of sign language classifiers, represent the size and shape of an object, such as 'cylindrical objects', 'flat objects', or 'small round objects', much as the Digueño examples do in (12). These handshapes may combine with motion morphemes (e.g., 'straight' or 'pivot'), location morphemes, and manner of movement morphemes, each with a meaning of its own. The classifier system is used to represent the type of referent engaged in an activity, its location, and the direction, shape, and manner of motion it manifests. The different combinations of elements of each type contribute to a system in which expressions can become very complex morphologically. In the examples below from ISL, classifiers for 'cylindrical object' and 'flat object' or 'vehicle' are combined with locations and motions to form complex structures. In the examples, each hand represents a different classifier. Figure (18.6a) means that a cylindrical object is placed next to a flat object, e.g., 'A cup is next to a piece of paper.' Figure (18.6b) means that two vehicles approached from opposite directions and passed each other.

A classifier may correspond to any of a long list of regular words. For example, the word for CAR is shown in Figure (18.6c). Objectively speaking, the phonetic forms of Figure (18.6b) and Figure (18.6c) are of roughly equal **phonetic** complexity. But the **morphological** complexity of each form is very different. Phonetically, both forms involve two hands with the same handshape, moving in opposition to one another, either by approaching and moving in opposite directions as in Figure (18.6b), or by moving in up-and-down arc shapes alternately, one hand after the other, as in Figure (18.6c). Yet morphologically, Figure (18.6b) is quite complex, involving two entities as well as their locations and movements relative to one another, while Figure (18.6c) is morphologically simple, meaning just CAR. Classifiers and lexical words have different uses and privileges of occurrence in the grammars of sign languages as well.[4]

It is very important to note that classifier predicates are linguistic entities, i.e., morphologically complex forms. They are neither pantomime nor otherwise simply analogic to real world things and activities. Furthermore, as the Digueño example shows, the type of morphology which incorporates nouns, verbs, and other lexical categories into single expressions is not uncommon in the world's spoken languages, though there are ways in which they differ, for example, in gradient and analogic structure influenced by iconicity in sign languages (see Emmorey (ed.) 2003).

The properties we have described at each level of grammatical structure – syntax, phonology, and morphology – provide strong evidence that certain basic characteristics of language

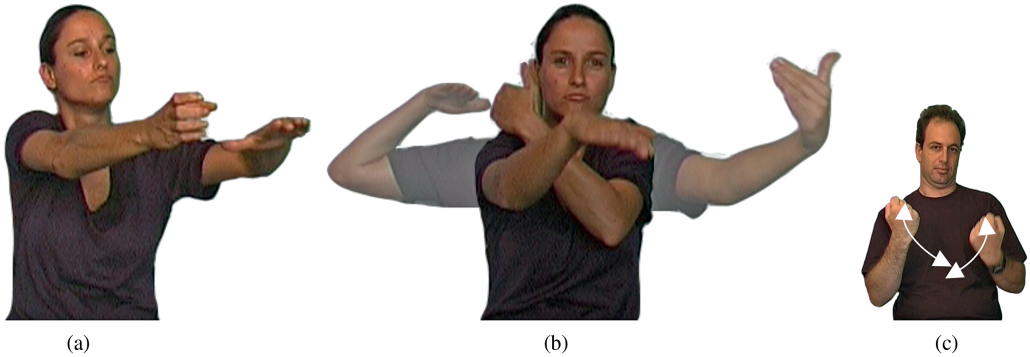


Figure 18.6 (a) 'A cup is next to a piece of paper.' (b) 'Two cars pass each other.' (c) CAR. © Wendy Sandler and Diane Lillo-Martin.

are indeed universal, belonging to language in either of the two physical modalities available to humans. Now we turn to consider other properties of sign languages which reinforce this conclusion, starting with sign language poetry.

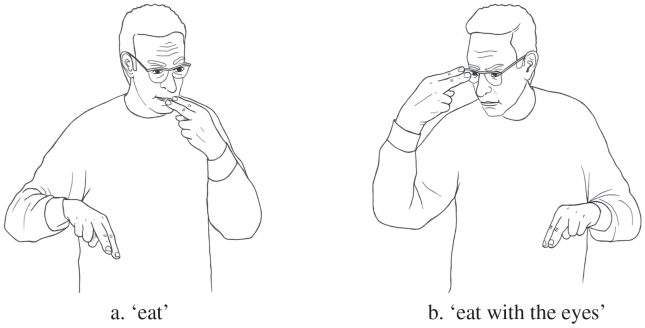
2 Language as an Art Form: Sign Language Poetry

Poetry takes language far beyond its primary task of everyday communication. By artfully manipulating the forms and meanings of language, the poet conveys a particular or heightened understanding of human experience. Devices such as meter, rhyme, and alliteration may filter the meaningful content of a poem, in order to create an impression or focus an image in the mind of the audience. A conventional meaning may be intentionally distorted in such a way as to enhance the perspective the poet wishes to present.

This union of language, culture, and art is found in some signing communities. We know of several accomplished deaf poets in different countries of the world, and offer as an example of sign language poetry some work of the late Wim Emmerik from Amsterdam.[5]

Among the devices used by Emmerik and observed in poetry of other sign languages as well are reiterative use of handshape, and a fluidity of style that results from the elimination of transitional movements. Entire poems may be characterized by one or two basic handshapes, such as the extended index finger handshape or a shape that extends all fingers . The aesthetic effect is similar to that of rhyme or alliteration in spoken poetry. While the movements that are part of signs are retained or modulated in some way, lines of poetry are skillfully constructed so as to omit the movements that result from the transition between the end of one sign and the beginning of another, creating flowing verse.

The explicitly poetic, somewhat paradoxical device of disrupting or distorting conventional meanings in order to enhance the intended meaning is also recruited by Emmerik. In his poem, *Member of Parliament*, Emmerik presents a picture of the governmental representative as jaded and cut off from the occurrences and sensations of the real world. The closest that the member of parliament gets to events in his country is through reading the newspaper as he eats lunch. To convey the idea that the politician crudely ingests the news through his body, rather than experiencing events spiritually or intellectually, the poet portrays him **eating** the news. As the politician eats and reads, Emmerik alters the usual sign for *eat* (Figure (18.7a)) by changing its location to the eyes rather than the mouth (Figure (18.7b)). Intermittently, the member of parliament eats food and ingests news of the world.



a. 'eat' b. 'eat with the eyes'

Figure 18.7 Sign language poetic form. © Wendy Sandler and Diane Lillo-Martin.

The value of the work of Emmerik and other deaf poets is first and foremost artistic. But the fact that poetry arises in established deaf communities is instructive as well. It teaches us that the desire and ability to mold the formational elements of language into an artistic expression of its content is a hallmark of human culture, whether the language is perceived by sound or by sight.

3 The Acquisition of Sign Languages

Research on the acquisition of sign languages addresses questions concerning the universality of proposed developmental milestones. Since sign languages and spoken languages share many linguistic characteristics, it might be expected that they would be acquired on the same maturational timetable. However, since sign languages rely largely on the development of relatively large manual articulators, while spoken languages rely on the more finely detailed use of the articulators of the vocal tract, there are possible differences attributable to the modality. If there were other, deeper differences between the development of signed and spoken languages this might call into question theoretical assumptions concerning the mechanisms for language development.

Numerous studies have examined the acquisition of sign languages by deaf children whose parents are deaf signers.[6] Although this population represents only about 5 percent of deaf children, it is the relevant one to study first because these children receive input in sign language from birth, just as hearing children receive input in their native language. In general, it has been found that deaf children who are exposed to sign language acquire it in very similar ways to the way hearing children acquire spoken languages: they pass through the same milestones at the same rate.[7]

As one example, let us consider the child's first words or signs.[8] Research on the acquisition of spoken languages finds that on the average, children begin to use their first words at around 11 months of age. When do the first words of sign language appear? Some early studies claimed that deaf children begin to use their first signs at around 6 to 8 months of age, much younger than the reported first spoken words. Why would there be such a difference between first words and first signs?[9] Two types of mutually compatible explanation have been proposed.

The first explanation points out that both deaf and hearing children make meaningful gestures during the time under consideration. Some researchers may attribute "word/sign" status to gestures made by signing children which are actually the same as gestures produced by speaking children and not considered words.

The second explanation considers the development of the articulatory mechanisms used for speech versus sign. Earlier maturation of the mechanisms for muscular coordination of the hands and arms over the vocal organs may make it easier for a child who is cognitively ready for first words/signs to produce a sign than a spoken word. Such explanations put the possible difference between the onset of signs and words in "peripheral" mechanisms rather than the cognitive timetable for language. The apparent discrepancy between the onset of first words in spoken versus signed language thus focuses important research questions that never would have arisen had we considered only spoken language.

Explanations which downplay the apparent difference between sign and spoken language development are all the more reasonable when other areas of language acquisition are considered, where there does not appear to be a difference between sign and speech in the attainment of each milestone. For example, another important milestone of language acquisition, the use of two-word(/sign) combinations, seems to be attained at around 18 months for both signing and speaking children. Other later grammatical developments, such as the acquisition of sentence structure, also follow parallel time courses across the modalities.

Overall, research emphasizes the similarity in acquisition between signed and spoken languages. Such a result strengthens the conclusion, already well grounded in theoretical research,

that sign languages are real languages, with all the scientific and social implications that follow from that conclusion.

Studies of the acquisition of sign languages are now able to go beyond this initial goal of establishing similarities (and accounting for differences) to use acquisition data for testing theories of particular grammatical structures. Likewise, developments in theoretical proposals concerning sign language structure can enlighten acquisition studies (Lillo-Martin 2016).

An example of this reciprocal relationship between acquisition studies and research on the (adult) syntax of a sign language comes from a pair of studies of word order in the Sign Language of the Netherlands (SLN). Coerts and Mills (1994) studied development of word order in SLN for two deaf children. The children showed a great deal of unexplained variability in their ordering of subjects and verbs. Subsequently, Bos (1995) showed that SLN has the process which Padden (1988), working on ASL, called Subject Pronoun Copy (SPC), as discussed in Section 18.1. Since SLN (like ASL) also allows the sentence-initial subject to be left unexpressed, pronominal subjects can end up appearing either before or after the verb, as illustrated in example (13) (from Coerts 2000).

(13) a. $\text{INDEX}_{\text{beppie}}$ FILM $\text{INDEX}_{\text{beppie}}$
'Beppie is filming.'

b. CRY $\text{INDEX}_{\text{dolls}}$
'The dolls are crying.'

With Bos's proposal in mind, Coerts (2000) reanalyzed the child data previously studied. She employed a fairly strict criterion for acquisition of the SPC process, and found that the majority of the previously "unexplained" word order examples were in fact explainable as instances of this process. Coerts concludes:

> knowledge of the adult language steers the choice of analysis procedures used for acquisition data … an analysis procedure that takes subject pronoun copy into account results in a much clearer picture with respect to the acquisition of subject and verb position (Coerts 2000)

Not only do developments in syntactic analysis make for improved studies of acquisition, but acquisition studies can also provide evidence bearing on competing syntactic analyses of the adult language. Lillo-Martin and Quadros (2005) took this approach in their study of the acquisition of focus constructions in ASL and Brazilian Sign Language (LSB). As is the case for many languages, these two sign languages have more than one mechanism for placing the focus on particular parts of a sentence. One analysis proposes that two different focus structures are explained by the same syntactic mechanism: placing focused elements in the sentence-final position and doubling, a construction in which focused elements appear both in their sentence-internal position and the sentence-final position. This is illustrated in (14), where the word in parentheses may optionally occur with its double in focus position.

(14) a. JOHN (CAN) READ CAN
'John really CAN read.'

b. MARY (FINISH) GO BRAZIL FINISH
'Mary ALREADY went to Brazil.'

Lillo-Martin and Quadros (2005) studied children's development of these structures, and found that the two were acquired at the same time. In contrast, a third type of focus construction was not acquired at the same age, indicating that the similarity was not due only to the pragmatic function of focus. This result in the study of acquisition was taken to support the linguistic

analyses of the adult languages which posit the same mechanisms for both types of focus over those which posit distinct structures.

These examples show how studies of the acquisition of sign languages, like studies of spoken language acquisition, contribute to the development and refinement of theoretical analyses of structure. Such studies also contribute in many ways to theories of the nature of language and language development. This last goal can be accomplished in even more interesting ways when studying sign languages because of the special sociocultural conditions under which sign languages are often acquired.

Lenneberg (1967) theorized that there may be a "critical period" during which exposure to language is required in order for the natural mechanisms used for its acquisition to be triggered. Learning language after this period seems to be qualitatively different from early language acquisition. A unique opportunity for investigating this critical or sensitive period hypothesis is presented by deaf children with hearing parents. Even in caring home environments, the age at which deaf children are exposed to sign language varies. Sign language input may begin as late as at the age of 5, when children enter school, or perhaps even later.[10] What, then, does sign language development in these children tell us about the critical period hypothesis?

Studies of the development of ASL in children with no linguistic input until after the age of 5 confirms that such a delay leads to a different pattern of acquisition (Berk 2003; Berk and Lillo-Martin, 2012). Although many aspects of language can be acquired and used fluently, some linguistic structures, such as verb agreement, show the evidence of delayed exposure at the beginning of language acquisition. In studies of adults who have used ASL for many years (e.g., Newport 1990) it has been found that the age at which exposure to language began is critical for the eventual acquisition of a complete grammar. Researchers find that even after 50 years of signing, people who were exposed to ASL only after the age of 5 systematically differ in their use of some of the complex mechanisms of ASL, as compared with similar-aged adults who acquired ASL from birth in signing homes. These later learners are fully expressive and use many of the same structures as the native learners. However, their use of complex morphology and syntax is often very different. These results support and refine Lenneberg's proposal that there is a sensitive period for language acquisition in a way which would not have been possible if not for the study of sign languages.

The study of the natural development of a new sign language offers a unique perspective on the human language capacity and the hypothesis that there is a sensitive period for language acquisition. Research conducted by Judy Kegl, Ann Senghas, and their colleagues (e.g., Kegl, Senghas, and Coppola 1999; Senghas and Coppola 2001; Senghas 2003), has been charting the development of a new sign language in Nicaragua, where deaf people had been isolated from one another until 1977. At that time a school was established for deaf children, and a communication system of signs immediately began to develop spontaneously. The older children had come to the school with only their home signs, a communication system that deaf children develop in the absence of a language model in order to communicate with their families (see Goldin-Meadow 2005). The system that developed among the first group of signers is not fully conventionalized, and it is relatively simple in structure. Since they had had no language model during the critical period, their communication with one another remained pidgin-like. Over time, however, as children as young as four began to come to the school and take the pidgin system of the older children as a model, a more systematic and conventionalized language began to emerge in these younger signers. The study of the development of this language shows that the human brain will create a communication system as soon as people congregate and have a need to communicate. Like creole studies in spoken language, it also shows that children have the capacity to further expand and regularize even relatively simple input and make it into a bona fide language. Presumably, the original path of development of many familiar sign languages that have been studied by linguists shares some similarities with that of Nicaraguan Sign Language, and documenting the history of NSL in real time gives us insight into the creation and development of

deaf community sign languages generally. Another type of new sign language, that which arises *de novo* in isolated communities with a high percentage of deaf people, has recently been discovered and raises new issues in research on language. We discuss this type of new sign language in Section 5.3.

4 Neural Control of Sign Languages

Neurolinguists attempt to gain insight into the nature of language by determining which areas of the brain control various components of the language system. The main type of data that set this field in motion was the performance of people who incurred brain damage, such as the localized damage resulting from strokes. Numerous studies have found that for hearing, speaking people damage to certain areas of the left cerebral hemisphere typically results in some type of language dysfunction. Language deficits that result from brain damage are called *aphasia*. There are different types of aphasia, in which damage to different areas of the brain result in different types of language deficits. It is assumed that if a given function *y* is lost after damage to a certain area A, then that area is involved in the control or processing of that function for people without brain damage (see Chapter 16). As with other realms of investigation, similarities between signed and spoken languages have been found here, as well as intriguing differences.

Although damage to areas of the left hemisphere results in aphasia in hearing subjects, damage to areas of the right hemisphere typically do not. On the other hand, damage to the right hemisphere frequently results in the loss of various types of spatial functioning. For example, patients with damage to the right hemisphere may be unable to draw a complete picture of a clock (typically neglecting the left side), or they may get lost in places that were very familiar to them before the brain damage.

It was shown in Section 1 that the use of space plays an important role in the grammar of sign languages. For example, verb agreement and verbs of motion and location as described there both involve spatial representation and manipulation. This situation, in which spatial cognition operations are directly related to linguistic cognition, prompts basic questions about brain organization with respect to spatial and linguistic functions in deaf signers. Several studies of deaf signers who have suffered brain damage have revealed patterns of language breakdown which emphasize the similarity rather than the difference between signed and spoken language in the mapping of both linguistic and spatial abilities within the brain.

Poizner, Klima, and Bellugi (1987) present case studies of six deaf patients with brain damage, and they show a striking difference between deaf patients with damage to the right versus left hemispheres. Like hearing patients, deaf signers who have left hemisphere brain damage have aphasia – in this case, aphasia for sign language. Some patients have very slow, awkward signing, like the speech of a "Broca's aphasic" (named after the French physician who first associated the linguistic breakdown with damage to a particular area of the brain). Others have signing which is more fluent, but which doesn't make sense, like a "Wernicke's aphasic" (a syndrome which results from damage to a different area of the left hemisphere). However, these patients have generally intact spatial cognitive abilities, such as drawing, face recognition, or localization of objects.

Dsigners who experience damage to the right hemisphere also have severe spatial deficits. They show left neglect, get lost in the hospital, and lose the ability to draw or show spatial relations, just like hearing patients with right hemisphere damage. However, the most important point is this: their signing is not dramatically impaired. They sign fluently and meaningfully, even using the "spatial" grammatical devices, although they show some comprehension deficits, a point we will return to in Section 5.2. Other studies of deaf aphasics confirmed that sign language is controlled primarily in the left hemisphere, where spoken language is controlled, rather than in the right hemisphere, where visuo-spatial abilities are controlled. Researchers interpret these results as implying that neural specialization for language is determined more by the type of cognitive

operations involved in organizing the linguistic system than by the physical channel of transmission. However, recent research reveals a more complex and nuanced picture (See Section 18.5.2).

The evidence for key properties of a human language capacity that transcend the physical modality of its expression is by now diverse and compelling. There is something about human cognition that converges on a complex and rich language system with particular formal and even neurological characteristics, even when the evolutionarily dominant channel for its transmission is not available.[11]

Yet this is still not the whole story. Differences between signed and spoken languages are far from trivial (Meier, Cormier, and Quinto-Pozos 2002; Sandler 2010, in press). New techniques, recent findings, and novel interpretations of existing results offer tough challenges to certain received views, and point the way toward important research in the future.

5 Some Recent Discoveries and Challenges

A context for the future investigation of the relationship between language and cognition is the existing fundamental dispute about the nature of the language faculty. At one extreme is the strong Chomskyan view that language is an "organ," innately specified, and computationally divorced from other aspects of cognition (e.g., Chomsky 1986). Other scholars argue that the experience of the child acquiring language plays a stronger role in determining language form. Jackendoff (1997, 2011) develops a theory according to which language is the outcome of a combination of factors, some specific to language (and possibly genetically specified), and others that are found across human cognition systems. We now turn to certain outstanding questions and recent discoveries about sign language that bear on this controversy. When they are seriously addressed, we believe the answers will lead us to a far deeper understanding of the language capacity than would have been possible without sign language research.

5.1 Basic, unexplained similarities among sign languages

We have argued that sign languages bear important similarities to spoken languages. But we have only hinted at how similar the grammatical structures of sign languages are to one another. This important generalization needs to be explained.

Using various grammatical criteria, linguists sometimes find it instructive to group languages into categories or types. These criteria may be applied at any level of analysis – syntax, morphology, or phonology. For example, some languages have primarily the Swahili type of morphology (see example 11), in which morphemes that mark grammatical properties like agreement are added to base words one after another. Others have the Digueño type (see example 12), where noun-like and verbal elements can be incorporated into a single form. In syntax, some languages may productively allow omission of the subject in tensed clauses while others prohibit it. Phonologically, in some languages several consonants are permitted to occur together before a vowel appears; in others, only one consonant may appear at the beginning of a syllable. The point is that a given spoken language may fall into one of any number of categories at each level of description.

As we have hinted in Section 1, in many ways sign languages form a single language type, and one to which no spoken language belongs. If this is the case, then some essential questions arise, for both cognitive psychology and for linguistics. In the following paragraphs, we will demonstrate some of the typological traits of sign languages.[12]

Let us begin with the relationship between the elements of form and meaning. In Figure (18.6b) ('two cars pass each other') we showed a complex sign, with several meaningful elements or morphemes. We pointed out that some spoken languages have equally complex forms, with substantively the same types of morphemes in them. But there are two important generalizations that we now wish to emphasize: (1) all established sign languages that we know of have

precisely this type of morphology (American Sign Language, Israeli Sign Language, British Sign Language, Danish Sign Language, Swedish Sign Language, Japanese Sign Language, etc.); and (2) despite the large number of meaning elements (morphemes) in signs of this type, they are all articulated essentially simultaneously, or within what may be viewed as a single syllable.

It's not difficult to see that the first generalization is not true of spoken languages. Languages like English, for example, have nothing even remotely like this sort of word structure. As for the second, the spoken language equivalent would be a language like Digueño in example (12) above, but in which all the different meaning elements were pronounced within a single syllable. Such a spoken language analogue – a language whose words contain many meaningful components within a single syllable – does not exist.

The sign language system is rule governed, grammatical, and violates no principles of universal grammar that we are aware of. It is also acquired by children at an age appropriate for acquisition of a complex morphological system. Yet the details of its form and use are particular to sign languages as a group.

Established sign languages generally also have the type of verb agreement we described in Section 1. The literature on the subject includes American Sign Language, British Sign Language, Taiwan Sign Language, Swedish Sign Language, Japanese Sign Language, Italian Sign Language, Danish Sign Language, Sign Language of the Netherlands, Israeli Sign Language, Brazilian Sign Language, and Australian Sign Language. In addition to formal similarity, sign languages apparently also share the division into classes of verbs which are arguably determined on the basis of meaning and aspects of form. Although these systems bear the imprint of typical sign language morphology, we hasten to add that they are not identical across sign languages. There are language-specific characteristics as well (see Meir 2002 ; Rathmann and Mathur 2002; Mathur and Rathmann 2012).

At the same time, the basic similarities are so robust that they emerge spontaneously in the contact language used among signers whose native sign languages are not mutually intelligible. Supalla and Webb (1995) studied deaf lecturers at international conferences, communicating to an audience of signers of over 20 different sign languages. In a contact language called International Sign,[13] these lecturers use the same kind of verb agreement that we have been describing here, in addition to other grammatical structures. The authors say that this is because these devices exist in the native sign languages of the lecturers – whatever those may be. We add that the signers may expect such devices to be understood because they are (intuitively) confident that these devices are sign language universal.

The similarities across sign languages in systems such as agreement contrast sharply with the situation in spoken languages: not all spoken languages have agreement, and the agreement systems of those that do may have certain different properties from language to language. Where there are classes of verbs with different agreement markers within a spoken language, these classes are arbitrary, and they are not determined by meaning. Nevertheless, we emphasize that the sign language systems in question are grammatical and rule governed; they violate no known universal principle of grammar, and they are acquired by children at the same age that similar types of structures are acquired in spoken languages.

In addition to these morphological similarities, sign languages all seem to share a particular type of phonological structure. For example, the constraint that only one finger group may characterize a morpheme applies to all sign languages we know of. Also, the fact that meaningful elements tend to be piled on simultaneously within single syllables, rather than linearly in a sequence, is a generalization about the phonological as well as the morphological structure of sign languages.

In short, sign languages form a language type. What makes this type different from the types posited for spoken languages is that the modality alone determines the type, and that no spoken language that we know of is of the sign language type. Rather, spoken languages fall into many different language types. Some researchers have suggested that the particular types of structures that sign languages tend to share may be common in the manual-visual modality because

they are iconically motivated, that is, motivated by a transparent correspondence between form and meaning (Aronoff, Meir, and Sandler 2005; Sandler and Lillo-Martin 2006; Wilbur 2008; Perniss et al. 2010; Meir, Padden, Aronoff, and Sandler 2013; Strickland et al. 2015; and Lepic, Börstell, Belsitzman, and Sandler 2016).[14] Aronoff et al. also point to language age as a factor, one which has received more attention recently with the documentation of nascent sign languages. We return to the issue of the relationship between language age and language structure in Section 5.3.

5.2 *Neurological differences*

Additional challenges are posed by more recent neurological findings. As reviewed in Section 4, aphasia studies show clearly that both spoken and signed languages are controlled primarily in the left hemisphere. More recent brain research on normal subjects using modern techniques such as Positron Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI) are suggestive of possible differences in brain organization for signed and spoken language. In particular, recent results find right hemisphere involvement in sign language processing by deaf and hearing native signers (see Corina and McBurney 2001, and Emmorey 2002 for overviews).

One research group used Regional Cerebral Blood Flow and Positron Emission Tomography to examine signed and spoken language representation in hearing native signers (hearing people born to deaf parents who learned sign language as a first language). The two techniques showed bilateral activation (activation in both cerebral hemispheres) for both modalities. Another research group, using Event Related Potentials, similarly found bilateral activation for both modalities, for both deaf and hearing native signers. Using functional Magnetic Resonance Imaging, they also found bilateral activation for processing ASL by deaf and hearing native signers; however, they did not find bilateral representation for English.

In these studies, the classical left hemisphere regions have been implicated for processing sign language, but in addition, some areas of the right hemisphere have also shown activity. Recall that patients with right hemisphere lesions reported in Section 4 also showed some comprehension deficits, indicating right hemisphere involvement in processing sign language. One study shows that the right hemisphere may be more involved in producing and comprehending certain topographic/spatial aspects of sign language, in particular those involving classifier constructions presented briefly in Section 1.3 (Emmorey et al. 2002). This result sits well with the known right hemisphere advantage for spatial processing. Findings of right hemisphere involvement for the spatially organized classifier system become even more interesting when added to discoveries of right hemisphere dominance for certain other functions in both spoken and signed language that may be related to the classifier system. In particular, some research has shown that processing words with imageable, concrete referents is controlled in the right hemisphere for speakers and for signers (Emmorey and Corina 1993). Findings such as these and the interpretations given them by researchers present a much more nuanced picture of the relationship between localization and cognitive functions, and they are another indication of the way in which sign language research adds important pieces to the puzzle of language organization in the brain.

5.3 *New sign languages in isolated village populations: the relationship between language age and language structure*

In their treatment of sign language morphology and the similarities found among established sign languages, mentioned in Section 5.1, Aronoff et al. (2005) also point out that other morphological complexities found in sign languages are more like those found in spoken languages. These patterns are specific to individual sign languages and they are arbitrary in nature, not

iconically motivated. The examples given by the authors are prefixes and suffixes that originate as independent words and become grammatical morphemes over time. For example, ASL has a negative suffix meaning "not at all" (that comes from the sign ZERO), while ISL also has a negative suffix (from the sign NOT-EXIST), that is different from the ASL suffix in form and meaning – it means '-less', as in the English word *clueless*. "Grammaticalization" processes of this sort, common in spoken languages, are still relatively rare in sign languages (but see Fischer and Lillo-Martin 1980; Pfau and Steinbach 2011). This should not be surprising, when the relatively young age of sign languages is taken into consideration. In spoken languages, grammaticalization processes can take hundreds of years to manifest themselves. Since no known sign languages are more than a few hundred years old, it makes sense that modality favored structure will be shared, while arbitrary encoding of grammatical properties characteristic of grammaticalization of affixes will only be sporadic and sparse in these languages. The authors predict that more devices that encode grammatical properties in a more arbitrary way are likely to accrue on a language-specific basis as sign languages age. This is one way in which sign language research illuminates the relationship between language age and language structure.

While well-studied sign languages that arose in deaf communities such as American Sign Language and British Sign Language are young, they have still been in existence for hundreds of years. But sign languages can arise anew in the span of a generation, and more recent discoveries are allowing researchers to observe even earlier stages in the life of a language. Recently, linguists and anthropologists have discovered a type of new sign language referred to as village sign languages. These languages typically arise in small, insulated villages in which a gene for deafness, combined with marriage within the group, may result in proportionally high numbers of deaf people, sometimes as high as 3.5 percent or 40 times higher than the average incidence elsewhere. In this situation, a sign language arises spontaneously, and is often used by deaf and hearing people alike. After the first generation, deaf children may have adult sign language models from birth. In recent years, village sign languages have been discovered in a Bedouin village in the Israeli Negev desert, on the north shore of Bali, Indonesia, in the foothills of the Akuapem mountains in eastern Ghana, and in other spots remote from one another and from national education systems.[15] Arising as they do in families within small communities with little or no influence from any other language, these fledgling languages may provide rare access to the quintessential ingredients of human language, and may even offer clues to the evolution of language in our species.

One question that arises is this: Are these languages born with the kind of complex linguistic system we have been describing here? That is, is grammatical complexity available as soon as people begin to communicate? Recent discoveries in the investigation of young village sign languages show that it takes time for a language to develop complex grammatical structure, and that the development of linguistic complexity may require particular social conditions as well. At the same time, the kernels of grammatical structure are there at the outset, as careful documentation of the course of their development has begun to reveal.

Some grammatical structure can arise very early, such as the robust Subject-Object-Verb word order found in second-generation signers of Al-Sayyid Bedouin Sign Language. This shows that distinguishing arguments of a verb – indicating who did what to whom – represent a linguistic property exploited by communicators at the outset (Sandler et al. 2005; Meir et al. 2017). But other mechanisms like verb agreement and classifier constructions that are typical of more mature sign languages apparently take longer to develop as systematic grammatical devices, though their roots can be discerned quite early (Aronoff et al. 2008; Sandler et al 2014). Compounding – putting two words together to make a new word – seems to be a process that arises early as a way of expanding the lexicon, and a particular type of compounding process that has not been reported in more familiar sign languages has been found in ABSL (Meir et al. 2010). And the basic property of phonology – duality of patterning – also arises gradually, showing that fully functional language can exist without it (Sandler et al. 2011).

While comparison among village sign languages is in the early stages (Zeshan and de Vos 2012; Meir et al. 2013), it appears from first indications that the amount and type of linguistic patterning varies from language to language, and that such factors as the size of a deaf community, its internal cohesion, and its age may influence the development of grammatical regularities. We hope that future research will test these hypotheses through rigorous description and comparison of village sign languages. The opportunity to study such possibilities evaded us as long as we restricted our study to spoken languages, all of which go back many millennia.[16]

It will also be of great interest to compare village sign languages to the young Nicaraguan Sign Language, developed among deaf children in a school setting, to other recently developed deaf community sign languages, such as Israeli Sign Language, and to the communication system of home signers, in order to measure the effects of the different social, linguistic, and communicative contexts on language structure. These studies are expected to contribute unique empirical data to academic discourse on issues such as language evolution, innateness, and the relationship between social factors and the development of language structure.

6 Conclusion

The study of sign language provides a natural laboratory for isolating certain fundamental properties of human language from the modality in which it is transmitted. Properties shared by languages in two such different physical modalities emerge as universal characteristics of human language. A meaningless, systematic subword level of structure (phonology); syntactic structure, embedding and recursion; prosodic marking of constituents and intonation; all are found in signed and spoken languages alike, and particular types of complex word formation found in many spoken languages are common in sign languages too. Sign language study has also strengthened the claim that the acquisition of language by children is a natural and automatic process with a set timetable, pointing to some degree of genetic predisposition for the development of just such a system.

At the same time, sign language research shows that neurological specialization is not identical for spoken and signed languages, that the physical modality in which a language is transmitted contribute to aspects of its grammatical system, and that the widespread occurrence of certain types of linguistic structure in sign languages does not necessarily mean that such complexity springs forth full-blown at the outset. These research directions are expected to make important contributions to contemporary thinking on such central topics as how the physical channel of transmission relates to language structure, which properties of language are universal in both modalities and what this tells us about our capacity for language, and the role of culture in the development of human language.

NOTES

1 Readers are referred to the following resources for comprehensive overviews of sign language research: *Sign Language and Linguistic Universals* (Sandler and Lillo-Martin 2006); *Sign languages* (Brentari 2010); *Sign Language: An International Handbook* (Pfau, Steinbach, and Woll 2012); and for a current bibliography, *Sign Language Linguistics* (Börstell, Aronoff, and Sandler 2015).

2 There has been considerable debate about what it would mean to find a language purported to lack embedding, in the context of reports about the Amazonian language Pirahã (see Everett 2009 and Nevins et al. 2009 for discussion and references to the works involved in this debate).

3 New sign languages may not have developed agreement systems; see Section 5.3.

4 The fact that the nondominant hand can be a separate morpheme in classifier constructions is one of the obvious ways in which these complex constructions differ from other signs, such as SHOW in Figure 18.5, in which each hand is involved only as a meaningless phonological element.
5 Sutton-Spence (2005) provides detailed analyses of sign language poetry.
6 See Newport and Meier (1985), Lillo-Martin (1999, 2016), Chamberlain, Morford, and Mayberry (eds) (2000), Chen Pichler (2012), and Chen Pichler et al. (2016) for reviews.
7 The process of language acquisition for deaf children with hearing parents depends on their linguistic input, which is quite varied. In the USA, Israel, and other countries, most deaf children now receive cochlear implants and therapy in spoken language, but they may also be exposed to some form of manual communication before and/or after implantation. A discussion of the pros and cons of early sign language use among deaf children can be found in Napoli et al. (2015). Although there are many interesting properties of language acquisition in differing sets of circumstances, we cannot go into them here.
8 See Meier and Newport (1990) for a thoughtful review of this research, on which this summary is based.
9 We acknowledge that determining the onset of first words requires studying a large number of children, because of the range of individual differences in the achievement of such milestones. Yet the findings so far indicate that signers may have an edge on speakers, and including these findings in the discussion illuminates the issue of potential modality differences.
10 Nowadays, many deaf youngsters in some countries are exposed to some form of signing at a relatively early age through school, community, or home programs which may begin when the child is little over a year old. However, even now many children are not exposed to sign language, or even any contrived sign system, until much later.
11 Presumably, if evolution had selected both modalities, there would be hearing communities that just happen to use a sign language for their primary means of communication. Since no such communities are known, we may assume that the evolutionary preference is for spoken language, whatever the reason.
12 In the discussion, we speak of "sign languages" or "all sign languages." What we mean by this is all mature, deaf community sign languages about which we have relevant information, including some sign languages not known to have any historical relationship to each other. We do not include village sign languages, which may lack some of these complexities.
13 International Sign is the form of communication used by deaf people with no common sign language, when they meet, for example at international conferences. It is not a contrived language, like Esperanto (or its sign language counterpart, Gestuno, neither of which has caught on), but rather arises naturally out of contact situations.
14 Some spoken language research suggests that iconicity plays a nontrivial role in spoken language structuring as well, though somewhat differently and to a more limited degree.
15 While the precise age of some village sign languages, such as Al-Sayyid Bedouin Sign Language, is known, the age of others, such as the sign language of Bali, is not certain.
16 Even pidgins and creoles, though newly formed, are descended from old languages.

REFERENCES

Aronoff, M., Meir, I., Padden, C., and Sandler, W. (2008). The roots of linguistic organization in a new language. *Interaction Studies: A Special Issue on Holophrasis Vs. Compositionality in the Emergence of Protolanguage*, 9(1), 131–50.

Aronoff, M., Meir, I., and Sandler, W. (2005). The paradox of sign language morphology. *Language*, 81(2), 301–34.

Berent, I. (2013). *The Phonological Mind*. Cambridge: Cambridge University Press.

Berk, S. (2003). *Sensitive Period Effects on the Acquisition of Language: A Study of Language Development*. PhD dissertation, University of Connecticut, Storrs.

Berk, S. and Lillo-Martin, D. (2012). The two-word stage: motivated by linguistic or cognitive constraints? *Cognitive Psychology*, 65, 118–40.

Börstell, C., Aronoff, M., and Sandler, W. (2015). Sign language linguistics. In M. Aronoff (ed.), *Oxford Bibliographies in Linguistics Online*.

Bos, H. (1995). Pronoun copy in Sign Language of the Netherlands. In H. Bos and T. Schermer (eds), *Sign Language Research 1994: Proceedings of*

the Fourth European Congress on Sign Language Research (pp. 121–47). Hamburg: Signum.

Brentari, D. (ed.) (2010). *Sign Languages: Cambridge Language Surveys*. New York: Cambridge University Press.

Brentari, D. (2011). Sign language phonology. In J. Goldsmith, J. Riggle, and A. Yu (eds), *Handbook of Phonological Theory, Second Edition* (pp. 691–721). New York/Oxford: Blackwell.

Chamberlain, C., Morford, J. P., and Mayberry, R. I. (eds) (2000). *Language Acquisition by Eye*. Mahwah, NJ: Lawrence Erlbaum.

Chen Pichler, D. (2012). Acquisition. In R. Pfau, M. Steinbach, and B. Woll (eds), *Sign Language: An International Handbook* (pp. 647–86). Berlin: Walter de Gruyter.

Chen Pichler, D., Kuntze, M., Lillo-Martin, D., Quadros, R. M. de, and Stumpf, M. R. (2016). *Sign Language Acquisition by Deaf and Hearing Children: A Bilingual Introductory Digital Course*. Washington, DC: Gallaudet University Press.

Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger

Coerts, J. (2000). Early sign combinations in the acquisition of Sign Language of the Netherlands: Evidence for language-specific features. In C. Chamberlain, J. Morford, and R. Mayberry (eds), *Language Acquisition by Eye* (pp. 91–109). Mahwah, NJ: Lawrence Erlbaum.

Coerts, J. and Mills, A. (1992). Early sign combinations of deaf children in Sign Language of the Netherlands. Paper presented at The Fifth International Symposium on Sign Language Research, Salamanca, Spain.

Corina, D. and McBurney, S. (2001). The neural representation of language in users of American Sign Language. *Journal of Communication Disorders*, 34(6), 455–71.

Corina, D. and Sandler, W. (1993). Phonological structure in sign language. *Phonology*, 10(2), 165–208.

Emmorey, K. (2002). *Language, Cognition, and the Brain: Insights from Sign Language Research*. Mahwah, NJ: Lawrence Erlbaum.

Emmorey, K. (ed.) (2003). *Perspectives on Classifier Constructions in Sign Languages*. Mahwah, NJ: Lawrence Erlbaum.

Emmorey K. and Corina, D. (1990). Lexical recognition in sign language: effects of phonetic structure and morphology. *Perception and Motor Skills*, 71, 1227–52.

Everett, D. (2009). Pirahã culture and grammar: a response to some criticisms. *Language*, 85(2), 405–42.

Fischer, S. D. and Gough, B. (1978). Verbs in American Sign Language. *Sign Language Studies*, 7, 17–48.

Fischer, S. D. and Lillo-Martin, D. (1990). UNDERSTANDing Conjunctions. *International Journal of Sign Linguistics*, 1, 71–80.

Goldin-Meadow, S. (2003). *The Resilience of Language: What Gesture Creation in Deaf Children Can Tell Us about How all Children Learn Language*. In the Essays in Developmental Psychology Series (J. Werker and H. Wellman, eds). New York: Psychology Press.

Goldin-Meadow, S. (2005). What language creation in the manual modality tells us about the foundations of language. *The Linguistic Review*, 22, 199–225.

Hockett, C. (1960). The origin of speech. *Scientific American*, 203: 88–111.

Jackendoff, R. (1997). *The Architecture of the Language Faculty.* Cambridge, MA: MIT Press.

Jackendoff, R. (2011). What is the human language faculty? two views. *Language*, 87(3), 586–624.

Kegl, J., Senghas, A., and Coppola, M. (1999). Creation through contact: sign language emergence and sign language change in Nicaragua. In M. DeGraff (ed.), *Language Creation and Language Change: Creolization, Diachrony, and Development* (pp. 197–237). Cambridge, MA: MIT Press.

Klima, E. and Bellugi, U. (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.

Langdon, M. (1970). *A Grammar of The Mesa Grande Dialect*. Berkeley and Los Angeles, CA: University of California Press.

Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley.

Lepic, R., Börstell, C., Belsitzman, G., and Sandler, W. (2016). Taking meaning in hand: iconic motivation in two-handed signs. *Sign Language and Linguistics*, 19(1), 37–81.

Liddell, S. K. (1980). *American Sign Language Syntax*. The Hague: Mouton.

Lillo-Martin, D. (1999). Modality effects and modularity in language acquisition: the acquisition of American Sign Language. In T. K. Bhatia and W. C. Ritchie (eds), *Handbook of Language Acquisition* (pp. 531–67). San Diego, CA: Academic Press.

Lillo-Martin, D. (2009). Sign language acquisition studies. In E. Bavin (ed.), *The Cambridge Handbook of Child Language* (pp. 399–415). Cambridge: Cambridge University Press.

Lillo-Martin, D. (2016). Sign language acquisition studies. In E. L. Bavin and L. R. Naigles (eds), *The Cambridge Handbook of Child Language*, 2nd edn (pp. 504–26). Cambridge: Cambridge University Press.

Lillo-Martin, D. and Meier, R. P. (2011). On the linguistic status of 'Agreement' in sign languages. Target article, *Theoretical Linguistics*, 37, 95–141.

Lillo-Martin, D. and Quadros, R. M. de (2005). The acquisition of focus constructions in American Sign Language and Língua de Sinais Brasileira. In A. Burgos, M. R. Clark-Cotton, and S. Ha (eds), *Proceedings of the 29th Annual Boston University Conference on Language Development* (pp. 365–75). Somerville, MA: Cascadilla Press.

Mandel, M. (1981). *Phonotactics and Morphophonology in American Sign Language*. PhD dissertation, University of California.

Mathur, G. and Rathmann, C. (2012). Verb agreement. In R. Pfau, M. Steinbach, and B. Woll (eds), *Sign Language: An International Handbook* (pp. 136–157). Berlin: De Gruyter Mouton.

Meier, R. P. (1991). Language acquisition by deaf children. *American Scientist*, 9, 60–70.

Meier, R. P., Cormier, K., and Quinto-Pozos, D. G. (eds) (2002). *Modality and Structure in Signed and Spoken Languages*. Cambridge: Cambridge University Press.

Meier, R. P. and Newport, E. L. (1990). Out of the hands of babes: on a possible sign advantage in language acquisition. *Language*, 66, 1–23.

Meir, I. (2002). A cross-modality perspective on verb agreement. *Natural Language and Linguistic Theory*, 20, 413–50.

Meir, I., Aronoff, M., Börstell, C., Hwang, S., Ilkbasaran, D., Kastner, I. … Sandler, W. (2017). The effect of being human and the basis of grammatical word order: insights from novel communication systems and young sign languages. *Cognition*, 158: 189–207.

Meir, I., Israel, A., Sandler, W., Padden, C., and Aronoff, M. (2012). The influence of community on language structure: evidence from two young sign languages. *Linguistic Variation*, 12(2), 247–91.

Meir, I., Padden, C., Aronoff, M., and Sandler, W. (2013). Competing iconicities in the structure of sign language. *Cognitive Linguistics* 24(2), 309–43.

Meir, I., Sandler, W., Padden, C., and Aronoff, M. (2010). Emerging sign languages. In M. Marschark and P. Spencer (eds), *Oxford Handbook of Deaf Studies, Language, and Education, Volume 2* (pp. 267–80). Oxford: Oxford University Press.

Napoli, D. J., Mellon, N. K., Niparko, J. K., Rathmann, C., Mathur, G. … Lantos, J. D. (2015). Should all deaf children learn sign language? *Pediatrics*, 136(1), 170–76.

Nespor, M. and Sandler, W. (1999). Prosody in Israeli Sign Language. *Language and Speech*, 42, 143–76.

Nevins, A., Pesetsky, D., and Rodrigues, C. (2009). Evidence and argumentation: a reply to Everett (2009). *Language*, 85, 671–81.

Newport, E. L. (1981). Constraints on structure: evidence from American Sign Language and language learning. In W. A. Collins (ed.), *Aspects of the Development of Competence. Minnesota Symposium on Child Psychology, Vol. 14*. Hillsdale, NJ: Lawrence Erlbaum.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, 14, 11–28.

Newport, E. L. and Meier, R. P. (1985). The acquisition of American Sign Language. In D. Slobin (ed.), *The Cross-Linguistic Study of Language Acquisition Vol. 1* (pp. 881–938). Hillsdale, NJ: Lawrence Erlbaum.

Padden, C. A. (1988). *Interaction of Morphology and Syntax in American Sign Language*. New York: Garland.

Perniss, P., Thompson, R. L., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, 1.

Pfau, R. and Steinbach, M. (2011). Grammaticalization in sign languages. In H. Narrog and B. Heine (eds), *The Oxford Handbook of Grammaticalization* (pp. 681–93). Oxford: Oxford University Press.

Pfau, R., Steinbach, M., and Woll, B. (2012). *Sign Language Linguistics Handbook*. Berlin: Mouton de Gruyter.

Poizner, H., Klima, E., and Bellugi, U. (1987). *What the Hands Reveal about the Brain*. Cambridge, MA: MIT Press.

Rathmann, C. and Mathur, G. (2002). Is verb agreement the same cross-modally? In R. P. Meier, K. Cormier, and D. G. Quinto-Pozos, *Modality and Structure in Signed and Spoken Languages* (pp. 370–404). Cambridge: Cambridge University Press.

Ross, J. (1967). *Constraints on Variables in Syntax*. PhD dissertation. MIT.

Sandler, W. (1989). *Phonological Representation of the Sign: and Nonlinearity in American Sign Language*. Dordrecht: Foris.

Sandler, W. (2010). The uniformity and diversity of language: evidence from sign language. Response to Evans and Levinson. *Lingua*, 120(12), 2727–32.

Sandler, W. (2011). Prosody and syntax in sign languages. *Transactions of the Philological Society*, 108(3): 298–328.

Sandler, W. (2012a). Phonological organization of sign languages. *Language and Linguistics Compass*, 6(2), 162–82.

Sandler, W. (2012b). Visual prosody. In R. Pfau, M. Steinbach, and B. Woll (eds), *Sign Language: An International Handbook* (pp. 55–76). Berlin: De Gruyter.

Sandler, W. (In press). The challenge of sign language phonology. *Annual Review of Linguistics.*

Sandler, W., Aronoff, M., Meir, I., and Padden, C. (2011). The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory*, 29, 503–43.

Sandler, W., Aronoff, M., Padden, C., and Meir, I. (2014). Language emergence. In J. Sindell, P. Kockelman, and N. Enfield (eds), *The Cambridge Handbook of Linguistic Anthropology* (pp. 250–84). Cambridge: Cambridge University Press.

Sandler, W. and Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.

Senghas, A. and Coppola, M. (2001). Children creating language: how Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science*, 12(4): 323–328.

Senghas, A. (2003). Intergenerational influence and ontogenetic development in the emergence of spatial grammar in Nicaraguan Sign Language. *Cognitive Development* 18(4), pp. 511-531.

Stokoe, W. C. (1960) *Sign Language Structure*. Silver Spring, MD: Linstok Press.

Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M., and Pfau, R. (2015). Event representations constrain the structure of language: sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences*, 112(19), 5968–73.

Supalla, T. (1986). The classifier system in American Sign Language. In C. Craig (ed.), *Noun Classes and Categorization* (pp. 181–214). Philadelphia, PA: John Benjamins..

Supalla, T. and Webb, R. (1995). The grammar of international sign: a new look at pidgin languages. In K. Emmorey and J. Reilly (eds), *Language, Gesture, and Space* (pp. 333–54). Hillsdale, NJ: Lawrence Erlbaum.

Sutton-Spence, R. (2005). *Analyzing Sign Language Poetry*. New York: Palgrave Macmillan.

Thompson, H. (1977). The lack of subordination in American Sign Language. In L. Friedman (ed.), *On the Other Hand: New Perspectives on American Sign Language*. New York: Academic Press.

Wilbur, R. B. (2000). Phonological and prosodic layering of non-manuals in American Sign Language. In K. Emmorey and H. Lane (eds), *The Signs of Language Revisited* (pp. 215–47). Mahwah, NJ: Lawrence Erlbaum.

Wilbur, R. B. (2008). Complex predicates involving events, time and aspect: is this why sign languages look so similar? In Josep Quer (ed.), *Signs of the Time: Selected Papers from TISLR 2004* (pp. 217–50). Hamburg: Signum Verlag.

Woll, B. (ed.) (2006). Sign language: history. In K. Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edn (pp. 307–310). Amsterdam: Elsevier Science.

Zeshan, U. and de Vos, C. (eds) (2012). Sign languages in village communities: anthropological and linguistic insights. *Sign Language Typology Series No. 4*. Berlin: De Gruyter Mouton & Nijmegen: Ishara Press.

19 First Language Acquisition

BRIAN MACWHINNEY

The transition from babbling to the first words marks a major turning point in the life of the infant. With the production of the first words, the child makes a significant step toward integration with human society. The word "infant" derives from the Latin, meaning "unable to speak." This reflects the fact that, at the time of the first words, the infant ceases being an infant and becomes a child. Some societies believe that, until children begin to speak, they have not yet taken on their full human souls. For this reason, they only give children names after they begin to talk.

Fortunately, the ability to acquire language is present in almost every human child (Lenneberg 1967). Children who are born blind have few problems learning to speak, although they may occasionally be confused about words for colors or spatial locations. Children who are born deaf readily acquire a rich system of signs. If their parents do not know sign language, then they create a set of signs through a process of mutual negotiation with their parents. Even Helen Keller, who had lost both her hearing and sight, was still able to acquire language through symbols expressed in touch and motion. Children with neurological disorders, such as brain lesions or hydrocephalus, often acquire complete control over spoken language, despite a few months of early delay. Given the pervasiveness and inevitability of first language acquisition, we often tend to take the process of language learning for granted. But language is the most complex skill that a human being can master. The fact that nearly all of us succeed in this task indicates how remarkably well the structure of language adapts to our underlying abilities. Language is immensely complex, but its complexity is accessible to all of us.

1 Learning Sounds

1.1 Auditory processing and memory

Language learning begins in the womb. Here, the fetus can pick up the rhythm of the mother's voice and the overall cadence of human language. However, in the womb, the amniotic fluid muffles the sounds available to the fetus. When the baby is born, the auditory and visual world changes suddenly. As the amniotic fluid drains out of the ears and the child opens her eyes, she begins to hear sounds and see sights that were never present before. William James described the world of the newborn as a "booming, buzzing confusion." It is certainly true that the change from the world of the womb to the world outside the womb is radical and severe. But this does not mean that the child is totally unable to structure this new perceptual world.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

Auditory processing relies on extensive pre-processing of signals for pitch and intensity in the cochlea and the auditory nerve. The cochlea is organized much like a xylophone, with hair cells responding to low pitches at the outside and cells responding to high pitches at the other end of the spiral on the inside. This tonotopic organization for pitch and intensity is preserved when the sound reaches the auditory cortex. In order to distinguish between sounds like the /p/ in *pit* and the /b/ in *bit*, the auditory cortex must be able to track things like the onset of resonant sound patterns and the direction of change within specific concentrations of sound pitches called formants. In the 1970s, researchers discovered that human infants were specifically adapted at birth to perceive many of these contrasts. This gave rise to the belief that humans had developed some unique methods for categorically based sound detection. However, subsequent research showed that even chinchillas are capable of making many of these distinctions (Werker 1995). Thus, it appears that much of the basic structure of the auditory world can be attributed to fundamental processes in the mammalian ear.

Beyond this basic level of auditory processing, infants have a remarkable capacity to record and store sequences of auditory events. For example, if the six-month-old hears a sound pattern such as /badigudibagadigudigagidu/ repeated many times, the parts that are repeated will stand out and affect later listening. In this example, the repeated string is /digudi/. If the infant is trained on these strings, she will come to prefer to listen to new sound strings rather than to those that have the old /digudi/ string (Saffran, Aslin, and Newport 1996). This indicates that the infant has come to store the /digudi/ string as a statistically predictable language unit. Researchers believe that this form of statistical learning operates automatically whenever the child is attending to speech. During the first year, the child is exposed to several thousand hours of human language. By continually attending to the auditory patterns of her language, the child can build up a rich repertoire of expectations about the forms of words. However, during this early period, the child still has no idea about the link between sounds and meanings. From the infant's point of view, language is still nothing more than an entertaining, but rather superficial, experience.

In addition to demonstrating early abilities to store sequences of sounds, babies also demonstrate preferences for the language that resembles the speech of their mothers. Thus, a French infant will prefer to listen to French, whereas a Polish infant will prefer to listen to Polish. In addition, babies demonstrate a preference for their own mother's voice, as opposed to that of other women. Together, these abilities and preferences suggest that, during the first eight months, the child is learning a lot about the sounds of language. Although the child is not yet learning words, she is acquiring the basic auditory and intonational patterns of her native language. As she sharpens her ability to hear the contrasts of her native language, she begins to lose the ability to hear contrasts not represented in her native language. If the child is growing up in a bilingual world, full perceptual flexibility is maintained. However, if the child is growing up monolingual, flexibility in processing is gradually traded off for quickness and automaticity.

1.2 Early articulation

Our understanding of infants' auditory processing is based on inferences made from experimental manipulations. In contrast, our understanding of articulatory development is based on a much richer array of directly observable behaviors. During the first three months, a baby's vocalizations can be characterized as various types of cries and vegetative adaptations (burping, sneezing, swallowing, etc.). At around three months, just after the time of the first social smiles, babies begin to make the delightful little sounds that we call "cooing." These sounds have no particular linguistic structure, but their well-integrated intonation makes them sure parent pleasers. The production of these sounds is supported by the infant's increased control of the larynx to produce stable phonations (Oller 2000). By six months, the baby is producing somewhat more structured vocalizations, including a larger diversity of nasals, vowel types, and syllables with the canonical consonant-vowel (CV) structure. The basic framework of early babbling can be

related to patterns of noisy lip-smacking that are present in many primate species (MacNeilage 1998). These vocal gestures include some form of vocal closure followed by a release with vocalic resonance. Essentially, this is the CV syllable in which a consonant is followed by a vowel.

Until the sixth month, deaf infants continue to babble normally. However, by the age of nine months, deaf infants have lost their interest in babbling. This suggests that the babbling present at six months is sustained largely through proprioceptive and somaesthetic feedback, as the baby explores the various ways in which she can play with her mouth. After six months, babbling relies increasingly on auditory feedback. During this period, the infant tries to produce specific sounds to match up with specific auditory impressions. It is at this point that the deaf child no longer finds babbling entertaining, since she cannot obtain auditory feedback. These facts suggest that, from the infant's point of view, babbling is essentially a process of self-entertainment.

Between six and ten months, there is a tight linkage between babbling and general motoric arousal. The child will move her arms, head, and legs while babbling, as if babbling is just another way of getting exercise while aroused. During the last months of the first year, the structure of babbling becomes clearer, more controlled, and more organized. Some children produce repetitive syllable strings, such as /badibadi badibadigu/; others seem to be playing around with intonation and the features of particular articulations.

In the heyday of behaviorism, researchers viewed the development of babbling in terms of reinforcement theory. They thought that the reinforcing qualities of language would lead a Chinese baby to babble the sounds of Chinese, whereas a Quechua baby would babble the sounds of Quechua. This was the theory of "babbling drift." However, closer observation of the babbling of eight-month-olds indicates that virtually no such drift occurs. By 12 months, there is some slight drift in the direction of the native language, as the infant begins to acquire the first words. Proponents of universal phonology have sometimes suggested that all children engage in babbling all the sounds of all the world's language. Here, again, the claim seems to be overstated. Although it is certainly true that some English-learning infants will produce Bantu clicks and Quechua implosives, not all children produce all of these sounds.

2 Learning Words

2.1 The first words

The child's ability to produce the first word is based on three earlier developments. The first is the infant's growing ability to record the sounds of words. The second is the development of an ability to control vocal productions that occurs in the late stages of babbling. The third is the general growth of the symbolic function, as represented in play, imitation, and object manipulation. Piaget characterized the infant's cognitive development in terms of the growth of representation through what he called the "object concept." In the first six months of life, the child is unable to think about objects that are not physically present. However, as the infant learns more about objects, she becomes able to associate their properties with her own actions and other features of the context. In this way, subtle cues can be used to dredge up fuller representations from memory. For example, a child may see a dog's tail sticking out from behind a chair and realize that the rest of the dog is hiding behind the chair. This understanding of how parts relate to wholes supports the child's first major use of the symbolic function. When playing with toys, the 12-month-old will begin to produce sounds such as *vroom* or *bambam* that represent enactive properties of these toys and actions. Often these phonologically consistent forms appear before the first real words. Because they have no clear conventional status, parents may tend to ignore these first symbolic attempts as nothing more than spurious productions or babbling.

If we look at early word learning from the viewpoint of the child, we realize the first steps toward learning words are taken in a fairly passive way. Even before the child has produced her

first conventional word, she has already acquired an ability to comprehend as many as ten conventional forms. She learns these forms through frequent associations between actions, objects, and words. Parents often realize that the prelinguistic infant is beginning to understand what they say. However, they are hard pressed to demonstrate this ability convincingly. Researchers deal with this problem by bringing infants into the laboratory, placing them into comfortable highchairs, and asking them to look at pictures, using the technique of visually reinforced preferential looking (Woodward, Markman, and Fitzsimmons 1994). A name such as "dog" is produced across loudspeakers. Pictures of two objects are then displayed. In this case, a dog may be on the screen to the right of the baby and a car may be on the screen to the left. If the child looks at the picture that matches the word, a toy bunny pops up and does an amusing drum roll. This convinces babies that they have chosen correctly and they then do the best they can to look at the correct picture on each trial. Some children get fussy after only a few trials, but others last for 20 trials or more at one sitting and provide reliable evidence that they have begun to understand a few basic words. Many children show this level of understanding by the tenth month (Oviatt 1980) – often two or three months before the child has produced a recognizable "first word."

This assessment may actually underestimate the time of the first auditory word. Even earlier, there is evidence that the child responds differentially to her own name. If two tapes are played to the right and left side of the six-month-old, the baby will tend to prefer to listen to the tape that includes her own name (Jusczyk 1997). Given the frequency with which the parent uses the child's name and the clarity with which it is typically presented, this is perhaps not too surprising. Although it is unclear whether the child actually realizes what this form means, she is clearly sensitive to its presence and responds when her name is produced.

Given the fact that the ten-month-old is already able to comprehend several words, why is the first recognizable conventional word not produced until several months later? From the viewpoint of the infant, producing the first word is a bit like stepping out on stage without having had sufficient time to rehearse. When she was babbling for her own entertainment, the only constraints the infant faced were ones arising from her own playfulness and interest. Now, when faced with the task of producing word forms, the articulation has to be extremely accurate and within conventional limits. Many of the child's first attempts to produce comprehensible words are so far away from the correct target that even the most supportive parent cannot divine the relation. Eventually, the child produces a clear articulation that makes some sense in context. The parent is amazed and smiles. The child is reinforced and the first word is officially christened.

But all is still not smooth sailing. The child still has no systematic method for going from auditory forms for words she knows to the corresponding articulatory forms. Earlier experience with babbling provides some guide, but now the linkage requires increased precision and control over difficult articulators such as the tongue and the lips. The many simplifications that the one-year-old introduces to adult phonology are well known to students of phonological development (Vihman and Croft 2007). Children tend to drop unstressed syllables, producing *hippopotamus* as *poma*. They repeat consonants, producing *water* as *wawa*. And they simplify and reduce consonant clusters, producing *tree* as *pee*. All of these phonological processes echo similar processes found in the historical development and dialectal variation of adult languages. What is different in child language is the fact that so many simplifications occur at once, making so many words difficult to recognize.

2.2 *Early semantics*

The salience of early articulatory limitations tends to mask other, more subtle, challenges facing the toddler. With only a few words to her name, there is no great danger that one word will be confused with another. However, as the toddler's inventory of words grows, the challenge of keeping these words apart also grows. The toddler is torn between two opposing strategies. On

the one hand, children often try to be conservative in their first uses of words. For example, a child may use the word *dog* to refer only to the family dog and not to any other dog. Or a child may use the word *car* to refer only to cars parked in the driveway and not cars in any other context. This tendency toward undergeneralization can only be detected if one takes careful note of the contexts in which a child avoids using a word. The flipside of this coin is the strategy of overgeneralization. It is extremely easy to detect overgeneralizations. If the child calls a tiger a *kitty*, this is clear evidence for overgeneralization. Of course, it is always possible that the children really meant to say something like, *that animal over there reminds me a lot of my kitty*. However, if the child intended this, they would be relying on nonstandard ideas about how words are used.

At first, both undergeneralization and overgeneralization are applied in a relatively uncontrolled fashion. The child's first applications of undergeneralization are unreasonably rigid and she soon learns that most words apply to a wide range of possible referents. At the same time, the child's first attempts at generalization are also often wildly overproductive. For example, a child may use the word *duck* first to refer to the duck, then to the picture of an eagle on the back of a coin, then to a lake where she once saw ducks, and finally to other bodies of water. These pleonastic extensions of forms across situations are fairly rare, but they provide interesting commentary regarding the thinking of the toddler when they do occur.

It would be fair to say that all children engage in both undergeneralization and overgeneralization of word meanings (Dromi 1987). At the same time, it is remarkable how accurate children are in their early guesses at the correct meanings of words. They quickly come to realize that words can be used across a variety of situations in addition to the original context in which they were used. This is fortunate, since reality never repeats itself. If a child thought that a word was limited to use in the original context, there would seldom be an opportunity to reuse a word. Instead, the child has to take each context and decide which aspects are likely to be generalizable for repeated uses of the word. But figuring out how to reuse words is not a trivial problem. In fact, scholars from Plato to Quine have considered the task of figuring out word meaning to be a major intellectual challenge. Quine (1960) illustrated the problem by imagining a scenario in which a hunter is out on safari with a native guide. Suddenly, the guide shouts *Gavagai*! The hunter, who does not know the native language, has to quickly infer the meaning of the word. Does it mean *shoot now*! or *there's a rhino* or perhaps even *it got away*? Without some additional cues regarding the likely meaning of the word, how can the hunter figure this out?

The problem facing the toddler is similar to that facing the hunter. Fortunately, the toddler has some good cues to rely on. Foremost among these cues is the parent's use of joint attention and shared eye gaze to establish common reference for objects and actions. If the father says *hippo* while holding a hippopotamus in his hand, the child can use the manual, visual, verbal, and proxemic cues to infer that the word *hippo* refers to the hippopotamus. A similar strategy works for the learning of the names of easily produced actions such as falling, running, or eating. It also works for social activities such as *bath* or *bye-bye*. The normal child probably understands the role of shared eye gaze even before learning the first words. At three months, children maintain constant shared eye gaze with their parents. In normal children, this contact maintains and deepens over time. For autistic children, contact is less stable and automatic. As a result, autistic children may be delayed in word learning and the general growth of communication.

The importance of shared reference is obvious to most parents. In fact, in the fanciful recollections in his *Confessions* (405 CE), St. Augustine outlined an analysis not very different from the one presented here:

> This I remember; and have since observed how I learned to speak. It was not that my elders taught me words (as, soon after, other learning) in any set method; but I, longing by cries and broken accents and various motions of my limbs to express my thoughts, that so I might have my will, and yet unable to express all I willed or to whom I willed, did myself, by the understanding which Thou, my God, gavest me, practice the sounds in my memory. When they named anything, and as they spoke turned

> towards it, I saw and remembered that they called what they would point out by the name they uttered. And that they meant this thing, and no other, was plain from the motion of their body, the natural language, as it were, of all nations, expressed by the countenance, glances of the eye, gestures of the limbs, and tones of the voice, indicating the affections of the mind as it pursues, possesses, rejects, or shuns. And thus by constantly hearing words, as they occurred in various sentences, I collected gradually for what they stood; and, having broken in my mouth to these signs, I thereby gave utterance to my will. Thus I exchanged with those about me these current signs of our wills, and so launched deeper into the stormy intercourse of human life, yet depending on parental authority and the beck of elders.

Shared reference is not the only cue the toddler uses to pick out the reference of words. She also uses the form of utterances to derive the meanings of new words. For example, if the toddler hears *here is a zav*, she knows that *zav* is a common noun. However, if she hears *here is Zav*, then she knows that *Zav* is either a proper noun or perhaps the name of a quantity (Katz, Baker, and Macnamara 1974). If she hears *I want some zav*, she knows that *zav* is a quantity and not a proper or common noun. Cues of this type can give a child a rough idea of the meaning of a new word. Other sentential frames can give an even more precise meaning. If the child hears *this is not green, it is chartreuse*, then it is clear that *chartreuse* is a color. If the child hears, *please don't cover it, just sprinkle it lightly*, then the child knows that *sprinkle* is a verb of the same general class as *cover*. The use of cues of this type leads to a fast, but shallow, mapping of new words to new meanings.

2.3 *Mutual exclusivity and competition*

Even the fullest set of syntactic cues and the clearest shared attention cannot prevent occasional confusion regarding word meanings. Some of the most difficult conflicts between words involve the use of multiple words for the same object. For example, a child may know the word *hippo* and hear her toy hippo referred to as a *toy*. Does this lead her to stop calling the toy a *hippo* and start calling it a *toy*? Probably it does not, although it may lead her to pay increased attention to the word *toy*. Some have suggested that children are prevented from making this type of error by the presence of a universal constraint called "mutual exclusivity." This constraint holds that each object can only have one name. If a child hears a second name for the old object, she can either reject the new name as wrong, or else find some distinction that disambiguates the new name from the old. If mutual exclusivity constrains word meaning, we would expect children to show a strong tendency toward the first solution – rejection. However, few children illustrate such a preference (Merriman and Bowman 1989). The problem with the rejection solution is that objects almost always have more than one name. For example, a *fork* is also *silverware* and a *dog* is also an *animal*. Linguistic structures expressing a wide variety of taxonomic and metonymic relations represent a fundamental and principled violation of the proposed mutual exclusivity constraint. The most consistent violations occur for bilingual children who learn that everything in their world must, by necessity, have at least two names. Mutual exclusivity is clearly not a basic property of natural language.

One reason why researchers have tended to devote so much attention to mutual exclusivity stems from the shape of the laboratory situation in which word learning is studied. The child is presented with a series of objects, some old and some new, given a word that is either old or new, and then asked to match up the word with an object. For example, the child may be given a teacup, a glass, and a demitasse. She already knows the words *cup* and *glass*. The experimenter asks her to *give me the demitasse*. She will then correctly infer that *demitasse* refers to the object for which she does not have a well-established name (Golinkoff et al. 1992). In this context, it makes sense to use the new name as the label for some new object.

Instead of thinking in terms of mutual exclusivity, the child appears to be thinking in terms of competition between words, with each word vying for a particular semantic niche (Merriman

1999). The child also thinks in terms of the pragmatics of mutual cooperation. When two words are in head-on conflict and no additional disambiguating cues are provided, it makes sense for the child to assume that the adult is being reasonable and using the new name for the new object. The child assumes that, like a cooperative parent, the experimenter knows that the child has words for cups and glasses, so it only makes sense that the new word is for the new object.

In the real world, competition forces the child to move meanings around so that they occupy the correct semantic niche. When the parent calls the toy hippo a *toy*, the child searches for something to disambiguate the two words. For example, the parent may say *can you give me another toy*? or *please clean up your toys*. In each case, *toy* refers not just to the hippo, but also potentially to many other toys. This allows the child to shift perspective and to understand the word toy in the framework of the shifted perspective. Consider the case of a rocking horse. This object may be called *toy*, *horsie*, or even *chair* depending on how it is being used at the moment. This flexible use of labeling is an important ingredient in language learning. By learning how to shift perspectives, children develop powerful tools for dealing with the competitions between words. In this way conflicts between meanings give rise to complex structures and cognitive flexibility.

2.4 Humpty-Dumpty and Whorf

In Lewis Carroll's *Through the Looking Glass*, Humpty-Dumpty chastises Alice for failing to take charge over the meanings of words. As he puts it, "When I use a word, it means just what I choose it to mean – neither more nor less." Undoubtedly, many children attempt to adopt this take-charge attitude toward language learning. The problem is that, without understanding conventional meanings, both children and Humpty-Dumpty could find themselves using words in ways that no one else would understand.

Children often have a rather fixed agenda of items to be expressed and would love to find simple ways of expressing each of those items. For example, many children want to learn words for finger, hand, ball, dog, bottle, Mommy, Daddy, and food. Most languages will oblige the child by providing words for these very basic concepts. However, once we leave the level of the first 20 words, all bets are off. Languages like Korean or Navajo require the child to learn verbs instead of nouns. Moreover, the verbs they learn focus more on position, shape, and containment than do verbs in English. For example, the verb *'ahééníshtiih* in Navajo refers to *carrying around in a circle any long straight object such as a gun.* As learning progresses, the child's agenda becomes less important than the shape of the resources provided by the language. This is not to say that languages end up shaping core features of children's cognitions. However, the presence of obligatory grammatical markings in languages for concepts such as tense, aspect, number, gender, and definiteness can orient the child's thinking in certain paths at the expense of others. Benjamin Whorf suggested many years ago that the forms of language shape the structure of thought. Such effects are directly opposed to the Humpty-Dumpty agenda-based approach to language. Probably the truth lies somewhere between Whorf and Humpty-Dumpty. Important though language-specific effects may be, all children end up being able to express basic ideas equally well, no matter what language they learn.

3 Learning Grammar

3.1 The first word combinations

Throughout the second year, the child struggles with perfecting the sounds and meanings of the first words. For several months, the child produces isolated single words. With a cooperative parent, a child can go a long way with this level of language. For example, if a child is hungry, it is enough to simply say *cookie*. There is no reason to say, *would you please open the cupboard door and*

bring me down a cookie. In fact, most of the child's basic needs are met even without the intervention of language. Sometimes a child may be frustrated by the parent's failure to understand her intentions. This frustration can be a strong motivator toward acquiring fuller communication. However, it is unlikely that needs and frustrations are the roots of linguistic development. Nor is language learned simply for the sake of imitating adults. Instead, it seems that children learn to speak so that they can articulate a fuller shared view of the world.

Single words are not enough to articulate this fuller view. Instead, children need to be able to associate verbs or "predicates" such as *want* or *go* with nouns or "arguments" such as *cookie* or *Mommy*. In addition to verbs, predicates can include other words that modify nouns such as adjectives, determiners and prepositions. In general, predicates are words that "say something" or "make a predication" about the nouns to which they are attached. This linkage of predicates to arguments is the first step in syntactic development (MacWhinney 2014). As in the other areas of language development, these first steps are taken in a very gradual fashion. Before producing a smooth combination of two words such as *my horsie*, children will often string together a series of single word utterances that appear to be searching out some syntactic form. For example, a child might say *my, that, that, horsie* with pauses between each word. Later, the pauses will be gone and the child will say *that horsie, my horsie*. This tentative combination of words involves groping on both intonational and semantic levels. On the one hand, the child has to figure out how to join words together smoothly in production. On the other hand, the child also has to figure out which words can meaningfully be combined with which others.

As was the case in the learning of single words, this learning is guided by earlier developments in comprehension. As in the case of studies of early word comprehension, we have to assess children's early syntactic comprehension by controlled experiments in the laboratory. Here, again, researchers have used the preferential looking paradigm. To the right of the child, there is a TV monitor with a movie of Big Bird tickling Cookie Monster. To the child's left, there is a TV monitor with a movie of Cookie Monster tickling Big Bird. The loudspeaker produces the sentence *Big Bird is tickling Cookie Monster*. If the child looks at the matching TV monitor, she is reinforced and a correct look is scored. Using this technique, researchers have found that 17-month-olds already have a good idea about the correct word order for English sentences. This is about five or six months before they begin to use word order systematically in production.

The level of successive single word utterances is one that chimpanzees also reach when they learn signed language. Domesticated chimps like Sarah, Washoe, or Kanzi have succeeded in learning over a hundred conventional signs or tokens. They can then combine these words to produce meaningful communication. However, the combinations that chimpanzees produce never really get beyond the stage of successive single word utterances. Thus, it appears that children rely on some uniquely human ability for structuring combinations of predicates and arguments into tighter syntactic combinations. The exact neurophysiological basis of this ability is still unknown, although many researchers suspect that the growth of inferior frontal areas for motor control supports the ability to put words into simple combinations.

The grammar of the child's first combinations is extremely basic. The child learns that each predicate should appear in a constant position vis-à-vis the arguments it requires. For example, in English, the word *more* appears before the noun it modifies. We can describe this relation as an item-based pattern in which the word *more* serves as the head item that opens up a slot that is filled by a following noun. In this case, *more* opens up a slot for a following noun. When a noun, such as *milk*, is selected to appear with *more*, that noun fills the slot opened up by the word *more*. The result is the combination *more milk*. Later, the child can treat this whole unit as an argument that drops into the direct object slot opened up by the verb *want* and the result is *want more milk*. Finally, the child can express the second argument of the verb *want* and the result is *I want more milk*. Thus, the child builds up longer and longer sentences and a more complex grammar. This level of simple combinatorial grammar is based on individual words as the controlling structures. This type of word-based learning is present even in adults. In languages with strong

morphological marking systems, word-based patterns specify the attachment of affixes, rather than just the linear position of words. In fact, most languages of the world make far more use of morphological marking than does English. In this regard, English is a rather exotic language.

3.2 Missing glue

The child's first sentences are almost all incomplete and ungrammatical. Instead of saying *this is Mommy's chair*, the child produces only *Mommy chair* with the possessive suffix, the demonstrative, and the copula verb all deleted. Just as the first words are full of phonological deletions and simplifications, the first sentences include only the most important words, without any of the glue. In some cases, children simply have not yet learned the missing words and devices. In other cases, they may know the "glue words" but find it difficult to coordinate the production of so many words in the correct order.

These early omissions provide evidence for two major processes in language development. First, the child makes sure that the most important and substantive parts of the communication are not omitted. Unfortunately, the child makes this evaluation from her own, egocentric, perspective. In an utterance like *Mommy chair* it is not clear whether the child means *this is Mommy's chair* or *Mommy is sitting in the chair*, although the choice between these interpretations may be clear in context. The second factor that shapes early omissions is phrasal frequency. Children tend to preserve frequent word combinations, such as *like it* or *want some*. These combinations are often treated as units, producing errors such as *I like it the ball* or *I want some a banana*.

In English, omissions of auxiliaries are extremely common. For many months, children will produce questions without auxiliaries, as in *why he go to the store*? for *why does he go to the store*? or *why not she come*? for *why won't she come*?" In languages with richer systems of morphological marking, the most common errors involve the use of the most frequent form of a noun or verb, even when some marked form is required. For example, in German child language, the infinitive is often used when a finite verb is required. These various errors can be traced to the fact that the child has limited resources to produce complex sentences and tends to settle for well-known forms in simple combinations.

3.3 Productivity

Productivity can be demonstrated in the laboratory by teaching children names for new objects. For example, we can show a child a picture of a funny looking creature and call it a *wug*. As we noted before, the positioning of the word *wug* after the article *a* induces the child to treat the word as a common noun. The child can then move from this fact to infer that the noun *wug* can pluralize as *wugs*, even if she has never heard the word *wugs* (Berko 1958). This type of productive generalization of linguistic patterns occurs from the earliest stages of language acquisition. For example, a German-speaking child can be taught the nonce name *der Gann* (nominative, masculine, singular) for a toy. The experiment can then pick up the toy and ask the child what he is holding. Even three-year-olds understand that *Gann* should be accusative. So, they correctly produce the form *den Gann*.

Three-year-olds also demonstrate some limited productive use of syntactic patterns for new verbs. However, children tend to be conservative and unsure about how to use verbs productively until about age five. After all, from the child's perspective these laboratory experiments with strange new toys and new words may tend to encourage a conservative approach. As they get older and braver, children start to show productive use of constructions such as the double object, the passive, or the causative. For example, an experimenter can introduce a new verb like *griff* in the frame *Tim griffed the ball to Frank* and the child will productively generalize to *Tim griffed Frank the ball*.

The control of productivity is based on two complementary sets of cues: semantics and co-occurrence. When the child hears *a wug*, she correctly infers that *wug* is a count noun. In fact, because she also sees a picture of a cute little animal, she infers that *wug* is a common, count, name for an animate creature. These semantic features allow her to generalize her knowledge by producing the form *wugs*. However, we could also view this extension as based on co-occurrence learning. The child learns that words that take the indefinite article also form plurals. On the other hand, words that take the quantifier *some* do not form plurals. In this way, the child can use both semantics and co-occurrence information to build up knowledge about the parts of speech. This knowledge about parts of speech can then be fed into existing syntactic generalizations to control the production of new combinations (Li, Zhao, and MacWhinney 2007).

3.4 The logical problem of language acquisition

The problem with productivity is that it produces overgeneralization. For example, an English-speaking child will soon learn to form the past tense of a new verb by adding one of the variant forms of *–ed*. This knowledge helps the child produce forms such as *jumped* or *wanted*. Unfortunately, it may also lead the child to produce an error such as **goed*. When this occurs, we can say that the child has formulated an overly general grammar. One way of convincing the child to reject the overly general grammar in which *goed* occurs is to provide the child with negative feedback. This requires the parent to tell the child, *no, you can't say "goed."* The problem here is that children may often ignore parental feedback regarding the form of language. If the child calls a hamburger a *hot dog*, the parent can tell her *no, it is a hamburger*. The child will accept this type of semantic correction. But children are notoriously resistant to being corrected for formal grammatical features.

The fact that children tend to ignore formal correction has important consequences for language acquisition theory. In the 1970s, work in formal analysis convinced some linguists that the task of learning the grammar of a language was impossible, unless negative feedback was provided. Since negative feedback appeared to be unavailable or unused, this meant that language could not be learned without some additional innate constraints. This argument has led to many hundreds of research articles exploring the ways in which children's learning places constraints on the form of grammar. Referring back to Plato's ideas about the difficulty of perceiving true forms, researchers have characterized the task of language learning as a logical problem. At its core, most of the search for innate constraints on language learning is grounded on the supposed impossibility of recovery from overgeneralization. To illustrate the ongoing importance of these issues for linguistic theory and language acquisition, consider this passage from Chomsky (1965: 58):

> It is for the present, impossible to formulate an assumption about initial, innate structure rich enough to account for the fact that grammatical knowledge is attained on the basis of the evidence available to the learner. Consequently, the empiricist effort to show how the assumptions about a language acquisition device can be reduced to a conceptual minimum is quite misplaced. The real problem is that of developing a hypothesis about initial structure that is sufficiently rich to account for acquisition of language, yet not so rich as to be inconsistent with the known diversity of language.

In fact, the child has more resources available to her than Chomsky seems to suggest. Using these resources, the child can recover from overgeneralization without negative feedback. In the case of *goed*, everyone agrees that recovery is easy. All the child has to do is to realize that there is only one way of producing the past tense of *go* and that is *went*. In other words, the irregular form *went* comes to block production of the over-regularized form *goed*. Here, recovery from overgeneralization is based on the competition between the regular pattern and the irregular form. In such competitions, the irregular form must always win.

However, not all recovery from overgeneralization is so simple. Suppose that a child decides to base the syntax of the verb *recommend* on that of the verb *give*. After all, both verbs involve a beneficiary and an object being transferred. However, only *give* allows a double object construction, as in *John gave the library the book*. Most people find *John recommended the library the book* ungrammatical. If the child makes this error, how does she recover? One solution to this error is to avoid making the error in the first place. If the child proceeds cautiously, learning each construction verb by verb, she will never attempt to use the verb *recommend* with the double object construction. Most children follow this course and never make the error. However, other children are less cautious. Do we want to assume that the cautious children have no need for innate constraints and that the less cautious children do? Fortunately there is a better way for even the incautious children to solve this "logical" problem.

The solution here is to record the strength of competing syntactic patterns. The correct way of saying *John recommended the library the book* is to say *John recommended the book to the library*. This correct formulation should be strengthened whenever it is heard. As the strength of the frame for the verb *recommend* grows in comparison to the ungrammatical frame, the use of the competing frame is blocked. This solution assumes that the child realizes that the two frames are in competition. It may be that reaching this realization requires some attention to syntactic form. However, this solution does not require the child to pay attention to corrective feedback. Instead, she only needs to attend to correct sentences and to make sure that she understands that these are competing ways of saying roughly the same thing.

3.5 *Lexical groups*

Most overgeneralizations can be controlled in a rote fashion. This involves strengthening single constructions for single verbs. However, there are some cases where stronger medicine may be necessary. Consider errors such as **I poured the tub with water* or **I filled water into the tub*. The use of a goal construction versus a direct object to express the entity being filled depends very much on the semantics of the verb. In effect, the child has to learn to break up the general class of pouring verbs into two subclasses, based on evidence from semantics and co-occurrence. Earlier, we discussed the role of lexical groups in supporting productivity. The problem here is the same one. However, the distinction is rather subtle, both semantically and syntactically. Verbs like *pour* do not emphasize the completion of the activity, but rather the ongoing process of transfer. These verbs use a goal construction. Verbs like *fill* tend to emphasize the completion of the activity and the change in state of the affected object. Most children learn to use these verbs conservatively and never produce these errors. However, once they are produced, the easiest way to correct them is to solidify the distinction between the two classes. Researchers (Li, Zhao, and MacWhinney 2007) have shown how the details of this learning process can be modeled formally using neural network models. Distinctions as subtle as this may not be acquired until the child produces some errors. Since errors of this type may not arise until about age six or later, the formation of lexical subclasses of this type is a rather late development.

Consider another example of how lexical classes help the child recover from overgeneralization. For example, a child might notice that both *cow* and *red* pattern together in forms such as *cow barn* and *red barn*. This might induce the child to produce forms such as *I painted the barn cow* on analogy with *I painted the barn red*. A conservative learner would stick close to facts about the verb *paint* and the arguments that it permits. If the child has heard a form like *I painted the barn white*, it would make sense to extend this frame slightly to include the resultative predicate *red*. However, an extension past the realm of colors and patterns would violate the basic principles of conservative learning. As a result, this type of category-leaping overgeneralization is extremely infrequent.

3.6 Errors that never occur

We have seen how children can recover from overgeneralization without relying on innate constraints. However, there is another approach to language development that provides more convincing evidence for innate constraints. This approach focuses on errors that "never" occur. Consider this example:

(1) a. The boy who is first in line will get the prize.
b. Will the boy who is first in line get the prize?
c. *Is the boy first in line will get the prize?

The claim here is that a simple surface analysis of the grammar would have led the child to produce (1c) instead of (1b). However, only (1b) is consistent with universal grammar, since auxiliaries are always derived from the main clause and not from some subordinate clause. Chomsky and others have claimed that children never hear sentences like (1b). It is certainly true that such sentences are not common, but it is not true that they never occur. Although the argument fails to go through in this case, the basic observation seems solid. Would a child even dream of producing something like (1c)? It seems unlikely. Moreover, it seems likely that, when the child learns to produce auxiliaries in questions, this learning is based not on surface word order, but on the underlying conceptual relations between words. It remains to be seen whether this learning amounts to evidence for innate constraints on grammar.

Similar analyses have been developed for a variety of other constructions. Examples (2) through (5) illustrate four additional patterns.

(2) a. You put it on a hot plate.
b. You put it on a hot what?
c. *What did you put it on a hot?

(3) a. Do you think a picture of Luke Skywalker should be on my cake?
b. Do you think a picture of who should be on my cake?
c. *What do you think a picture of should be on my cake?

(4) a. She chased the boy who stole her sandwich.
b. She chased the boy who stole her what?
c. *What did she chase the boy who stole?

(5) a. Luisa stood between the bookshelf and the fireplace.
b. Luisa stood between the bookshelf and what?
c. *What did Luisa stand between the bookshelf and?

In the case of (2c) and (3c), there is evidence that children actually produce the "nonoccurring" error. In fact, Bob Wilson's son Seth produced (2c) and my son Mark produced (3c). The corpora of child language data from which these examples were taken can be found in the CHILDES corpus on the web at http://childes.psy.cmu.edu.

Errors such as (4c) and (5c) have never been reported. Indeed, the constraints that block (4c) and (5c) are some of the most powerful constraints that have been identified in the linguistic literature. Both (4b) and (5b) seem to be possible ways of expressing these questions. However, they only make sense if we imagine conditions of noise that blocked out a single word. Not hearing well, we then echoed the sentence to try to recover the missing word. This suggests that neither (4c) nor (5c) is really well formed on semantic grounds.

3.7 Emergentist accounts

Our overview of language learning has focused on the challenges facing the child. We have also looked at language development from the viewpoint of Universal Grammar. Now we turn our attention to psychological views on language learning. Typically, psychologists see linguistic knowledge as emerging from regularities in the language heard by the child. To model the processes and mechanisms involved in this learning, many psychologists rely on the formalisms of neural network theory, which is also known as connectionism. This framework uses large numbers of units and the connections between these units to capture the patterns of language. This weblike architecture of nodes and connections is intended explicitly to resemble the structure of the human brain with neurons, synapses, and weights on synaptic connections.

Without burdening the reader with all the technical paraphernalia of neural network theory, let us take a brief look at how this type of analysis can be applied to a concrete problem in language acquisition. Let us take as an example the learning of German gender, as marked by the definite article (the word *the* in English). The task facing the German child is to combine each noun with one of the six different forms of the definite article. The article can take the form *der*, *die*, *das*, *des*, *dem*, or *den.* The choice of one of these six forms depends on three features of the noun: its gender (masculine, feminine, or neuter); its number (singular or plural); and its role within the sentence (subject, possessor, direct object, prepositional object, or indirect object). To make matters worse, assignment of nouns to gender categories is often quite nonintuitive. For example, the word for *fork* is feminine, the word for *spoon* is masculine, and the word for *knife* is neuter. Acquiring this system of arbitrary gender assignments is particularly difficult for adult second language learners. In his treatise on the "Aweful German Language," Mark Twain complained that German treats pretty young girls as neuter, the sun as feminine, and the moon as masculine. Twain was convinced that the choice of gender in German made no sense at all.

Although the cues governing German gender are complex, it is possible to construct a connectionist network that learns the German system from the available cues (MacWhinney et al. 1989). To do this, the network is presented with a series of patterns across the "input units." Each pattern represents the phonological form of a given German noun. For example, a particular node may be used to code the fact that the first consonant in the third syllable is a voiceless consonant like /p/ or /t/. Using 168 of these feature units, it is possible to given a different input pattern for each of the 102 nouns that were used to train the network. For each noun, the input also includes features that determine the noun's case and number.

Processing begins when the input layer is given a particular German noun. For example, the input could be the phonological form of the masculine noun *Tisch* (table), along with information that the noun is in the accusative and is singular. These active input units then spread activation to the other units in the system and eventually the activation reaches the six possible output units – one for each of the six forms of the definite article. The output unit that receives the most activation is the one that is chosen for the noun on this trial. On the first pass through, the network will probably choose the wrong output. In this case, the output might be the article *die*. This is wrong, since it treats the masculine noun *Tisch* as if it were feminine. When this occurs, the learning algorithm goes through all the connections in the network and adjusts them so that they are a bit closer to what would have been needed to activate the correct output item. This training continues for 50 cycles that repeat each of the nouns in the input corpus. At the end of this training period, the network is able to choose the correct article for 98 percent of the nouns in the original set.

To test its generalization abilities, we next present the network with old nouns in new case roles. If the network learned *Tisch* in the accusative, we now give it *Tisch* in the genitive and it should select the article *des.* In these tests, the network chooses the correct article on 92 percent of trials. This type of cross-paradigm generalization provides evidence that the network went beyond rote memorization during the training phase. In fact, the network quickly succeeds in

learning the whole of the basic formal paradigm for the marking of German case, number, and gender on the noun.

In addition, the network is able to generalize its internalized knowledge to solve the problem that had so perplexed Mark Twain – guessing at the gender of entirely novel nouns. The 48 most frequent nouns in German that had not been included in the original input set are then presented in a variety of sentence contexts. On this completely novel set, the network chooses the correct article from the six possibilities on 61 percent of trials, versus 17 percent expected by chance. Thus, the system's learning mechanism, together with its representation of the noun's phonological and semantic properties and the context, produced a good guess about what article would accompany a given noun, even when the noun was entirely unfamiliar.

The network's learning parallels children's learning in a number of ways. Like real German-speaking children, the network tends to overuse the articles that accompany feminine nouns. The reason for this is that the feminine forms of the article have a high frequency, because they are used both for feminines and for plurals of all genders. The simulation also showed the same type of overgeneralization patterns that are often interpreted as reflecting rule use when they occur in children's language. For example, although the noun *Kleid* (which means clothing) is neuter, the simulation used the initial "kl" sound of the noun to conclude that it is masculine. Because of this, it chooses the form of the definite article that would accompany the noun if it were masculine. Interestingly, the same article–noun combinations that are the most difficult for children are also the most difficult for the network.

How is the network able to produce such a high level of generalization and such rule-like behavior without any specific rules? The basic learning mechanism involves adjusting connection strengths between input, hidden, and output units to reflect the frequency with which combinations of features of nouns were associated with each article. Although no single feature can predict which article would be used, various complex combinations of phonological, semantic, and contextual cues allow accurate prediction of which articles should be chosen. This is the sense in which language learning often seems to be based on the acquisition of cues, rather than rules.

4 A Fourth Perspective

Alongside the perspective of the linguist, the psychologist, and the child, we can also look at language learning from the viewpoint of the parent and the educator. Parents often worry about the fact that their child may be slow at learning to talk. When a child falls behind, the parent and the educator want to know how to help the child catch up. However, experience shows us that the overwhelming majority of late talkers end up with full control over language. Often children are simply insufficiently motivated to talk. A prime example of this type is Albert Einstein, who did not begin talking until age five. His case is a bit extreme, but certainly not unique. Even children who have lost portions of their cerebral cortex as a result of early brain injuries end up acquiring full control over language use, as long as they are raised in a normal, supportive family.

Nearly 1 out of 20 children suffers from some form of language impairment. In many cases, language impairment is an accompaniment to some other obvious cognitive or emotional impairment, such as Down syndrome, Williams syndrome, Fragile-X syndrome, or autism. Each of these genetically based syndromes has a wide variance of expression, with some children achieving normal control of language and others less adequate language. Another, much larger, group of children evidences some level of language impairment without any obvious genetic abnormality. These children can be further divided into about four major groups. In the first group, only the expressive use of language is impaired. Children with expressive impairments may find it difficult to articulate certain sounds or may stutter. These children typically have

little impairment in language comprehension and no cognitive deficit. This deficit can be treated by articulatory speech therapy. A second group of children has deficits in low-level speech perception for sounds like /s/ and /f/. Careful training in the detection of auditory contrasts can help remediate this impairment. A third group of children shows some form of pragmatic impairment. These children have problems forming coherent discourse and connected narration. In some cases, this "deficit" may reflect stylistic effects related to dialect and social class. In other cases, it may reflect innate tendencies such as autism or difficulties with social perspective taking. Finally, there is a fourth group of children that have slight cognitive deficits that may be related to language impairments.

We are now just beginning to understand the neurological and genetic bases of these various impairments. Studies of familial genetic profiles have given us some clues regarding ways in which biology may determine language impairment. Recent advances in brain imaging methodology are now opening up the possibility of observing the actual neurophysiological bases of language processing as it occurs. Application of these new methods to the study of language impairments will help us better understand both normal and abnormal language development.

Not all parental concerns focus on language delay. Parents are also deeply interested in furthering normal progress and promoting genius. In some cases, the parent may find that the child has unusual interests in language and wants to help the child to develop these interests, whether they involve learning additional languages, growing up bilingual, or merely being introduced at an early age to great literature. Research on the roots of literacy has indicated the continuity between early literary practices such as reading books with children, reciting rhymes, or fantasy role-play and later success in reading and literacy (Snow 1999).

5 Conclusion

Language is a unique marker of humanity. It distinguishes the human species from the rest of the creation, and it allows us to share our thoughts and feelings. Language is the most complex skill that any of us will ever master. Despite this complexity, nearly every human child succeeds in learning language. This suggests that language is optimally shaped to mesh with the abilities of the human mind and body. On the one hand, the universals of human language match up with our neurological, cognitive, and physical abilities. At the same time, parents provide rich contextual and emotional support to guide children through the process of language discovery. By studying language learning, we learn more about universals of human language, the shape of social interaction, and the structure of the human mind.

SUGGESTIONS FOR FURTHER READING

Clark, E. (2009). *First Language Acquisition*. New York: Cambridge University Press.

De Houwer, A. (2009). *Bilingual First Language Acquisition*. New York: MM Textbooks.

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, 31, 883–914.

MacWhinney, B. (2005). Language evolution and human development. In B. Ellis and D. Bjorklund (eds) *Origins of the Social Mind* (pp. 383–410). New York: Guilford.

O'Grady, W. (2005). *Syntactic Carpentry*. Mahwah, NJ: Erlbaum.

Tomasello, M. (2003). *Constructing A First Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

RELEVANT JOURNALS AND SOCIETIES

Journals
Journal of Child Language
First Language
Language Learning and Development
Language Acquisition

Societies
Boston University Conference of Language Development
Child Language Seminar
International Association for the Study of Child Language
Stanford Child Language Research Forum

EMERGING TRENDS AND RESEARCH QUESTIONS

1. Is there any fundamental difference between first and second language acquisition, or are the differences a result of the fact that adults already have a language in place?
2. At age seven, many children are still disfluent, but over 80 percent of these children do not end up stuttering. Does this indicate that early intervention could help the other 20 percent?
3. To what degree can we attribute different patterns of academic success to early patterns of language usage and shared social routines? Which patterns lead to the best academic outcomes and how can they be encouraged in families and in the schools?
4. How can advances in the study of human and primate genetics help us understand the biological bases for the evolution of human language?

REFERENCES

Berko, J. (1958). The child's learning of English morphology. *Word*, 14, 150–77.
Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
Dromi, E. (1987). *Early Lexical Development*. New York: Cambridge University Press.
Golinkoff, R., Hirsh-Pasek, K., Bailey, L., and Wenger, N. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, 28, 99–108.
Jusczyk, P. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.
Katz, N., Baker, E., and Macnamara, J. (1974). What's in a name? A study of how children learn common and proper names. *Child Development*, 45, 469–73.
Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley.
Li, P., Zhao, X., and MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, 31, 581–612.
MacNeilage, P. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21, 499–546.
MacWhinney, B. (2014). Item-based patterns in early syntactic In T. Herbst, H.-J. Schmid, and S. Faulhaber (eds), *Constructions Collocations Patterns* (pp. 33–70). Berlin: de Gruyter Mouton.
MacWhinney, B., Leinbach, J., Taraban, R., and McDonald, J. L. (1989). Language learning: cues or rules? *Journal of Memory and Language*, 28, 255–77.
Merriman, W. (1999). Competition, attention, and young children's lexical processing. In B. MacWhinney (ed.), *The Emergence of Language* (pp. 331–58). Mahwah, NJ: Erlbaum.
Merriman, W. and Bowman, L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, 54, 3–4, 1–129.
Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Mahwah, NJ: Erlbaum.
Oviatt, S. (1980). The emerging ability to comprehend language: an experimental approach. *Child Development*, 51, 97–106.
Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA: MIT Press.
Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–8.
Snow, C. E. (1999). Social perspectives on the emergence of language. In B. MacWhinney (ed.), *The*

Emergence of Language (pp. 257–76). Mahwah, NJ: Erlbaum.

Vihman, M. and Croft, W. (2007). Phonological development: toward a "radical" templatic phonology. *Linguistics*, 45, 683–725.

Werker, J. (1995). Exploring developmental changes in cross-language speech perception. In L. Gleitman and M. Liberman (eds), *An Invitation to Cognitive Science. Language Volume 1* (pp. 87–106). Cambridge, MA: MIT Press.

Woodward, A. L., Markman, E. M., and Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, 30, 553–66.

Part V Languages in Use

20 Pragmatics

Language and Communication

RUTH KEMPSON

1 The Puzzle of Language Use: How Do We Ever Understand Each Other?

How language is used may not seem to warrant a topic on its own. "When *I* use a word, it means just what I choose it to mean – neither more, nor less." This is the view of language expressed by Humpty-Dumpty in Lewis Carroll's *Alice through the Looking Glass*. It may seem mere common sense that describing a person's knowledge of a language involves simply describing how they use the language. But this can't be *all* there is to say about the relation between our knowledge of language and the way we use it, for words invariably convey much more than they mean. The question is: What does this apparent gap between language use and meaning tell us about language? As Alice quite reasonably objected: "The question is whether you *can* make a word mean so many different things." The starting point for looking at language use is to consider why Humpty-Dumpty might have been right. Consider the conversation in (1):

(1) A: Can you cook?
B: I know how to put a kettle on.

Why are such conversations possible – why, for example, did B not just say "No"? After all, if B had understood the question, and knows what such a question "means," she should know that this type of question is a request for the answer *yes* or the answer *no*. So what did she "mean" by choosing that indirect mode of reply in apparent violation of this rule? In what sense could B have "meant" that she never did anything in the kitchen other than putting a kettle on, so that the answer is that she cannot cook, or has no interest in cooking, etc.

Then there is the way we can describe things by using words in ways which only have a very indirect relation to what a dictionary would indicate is their meaning:

(2) You're a real racehorse.[1]

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

How can the words *real racehorse* be predicated of a single individual, someone transparently not a horse? What could a speaker mean by saying something apparently blatantly far from the truth? One has sympathy with Humpty and yet also with Alice. What is it that we do with language that makes statements such as (2) meaningful and effective? What these two examples show is that words can be used to convey both more than what they conventionally mean and also something quite different. But, if this is true, how do we manage to sort out which is which? How do we know when an expression is to be taken at its face value, when it is to be taken as conveying rather more than what it actually presents, and when it has to be interpreted in some other, metaphorical, way? Indeed, what does this difference between these apparently distinct types of usage amount to? This is only the beginning of the problem of understanding meaning in language use, for even setting aside supposedly special rhetorical effects as in (2), expressions in language are notoriously prone to ambiguity. Yet, by and large, we have no difficulty in sorting out what the speaker is intending to say to us. How do we manage this? To take an extreme example (Sperber and Wilson 1982), consider (3):

(3) A: How is your son?
B: He's grown another foot.

Why is it that A is most unlikely to respond with horror, suggesting that it should be amputated? And, with a much less extreme case, how does the hearer decide whether the phrase "in March" describes the time of the exam, the time of discovering the results of the exam, or the time at which Sue made her report:

(4) Sue reported to the Committee that Joan learnt that she had failed the exam in March.

Even in (1) itself, B's reply could have been intended as a deliberate understatement indicating her cooking talent, which she may be implying A ought to know about. How does A decide which B meant? To put the problem in its most general form, when we probe the obvious truth that our knowledge of language is reflected in the way we use it, we seem to be faced with a perplexing mystery. How is it that using language is in general so effortless, when there seems to be no limit to what words can convey or what ambiguities they can give rise to? *Pragmatics* seeks to provide an answer to this question. Pragmatics is the study of communication – the study of how language is used. This study is based on the assumption of a division between knowledge of language and the way it is used[2]; and the goal of pragmatics is taken to be that of providing a set of principles which dictate how knowledge of language and general reasoning interact in the process of language understanding, to give rise to the various different kinds of effects which can be achieved in communication. In this chapter, we shall look at different approaches to pragmatics. We will look first at the question of the assumed interaction between language-particular (= linguistic) and general (= nonlinguistic) types of information; and it will be argued in Section 2 that what we need is a model which allows integration of linguistic and nonlinguistic forms of interpretation at all stages of the interpretation process. In Section 3 we will take up the question of how the hearer chooses an appropriate interpretation, and the two major alternative views of how interpretations are selected will be introduced. In Section 4, the view will be set out that all words should be defined in terms of procedures for building up structures representing interpretation in context (*propositional structure*). It will give one illustration to show how pragmatic processes can feed into linguistic processes, and then we shall see how these linguistic processes can freely interweave with pragmatic processes in conversational dialogue. The chapter ends with a discussion of the general significance of this "procedural" approach for establishing what it means to "know a language."

2 Pragmatics as the Application of Conversational Principles to Sentence Meanings

The starting point for studies in pragmatics[3] is the mismatch, often a big one, between what words "mean," which is encoded in rules of the language, and what speakers "mean" by using them, which may be much richer. In (1), the words in the sentence B utters convey the information that the speaker has the knowledge of how to put a kettle on. What speaker B means by using these words (on the interpretation indicated first) is that she cannot cook, opening up the possibility of further inferences such as that she has no interest in cooking. One way in which one might seek to generalize from this particular instance is to take (1) as evidence for the two aspects of language use being quite separate. On one standard view, there is the knowledge of language, on the one hand, which dictates the meanings of words and the ways in which they can combine to form sentence-meanings (to be studied under the label *semantics* as part of the grammar of a language). This is called the encoded meaning. On the other hand, there are general pragmatic principles (which I shall initially call "commonsense reasoning" principles) which enable a hearer to establish some rather different and richer interpretation – the nonencoded part of meaning. On this view, we would say that a hearer parsing B's utterance in (1) above first uses rules of the language to work out that B was conveying that she knew how to put a kettle on, and then, only subsequently, uses principles of commonsense reasoning to work out, say, that because B did not reply more directly, it must be that she is expecting A to access the knowledge that people who confess to knowing only how to put on a kettle are no good at cooking in order to work out the intended import of her answer – that she cannot cook. Once such general forms of reasoning are invoked, one might also anticipate that such indirect forms of answer convey more information than a simple negative answer would provide because they trigger such general reasoning processes – for example, communicating from B's implication that she has no interest in cooking the further suggestion that any attempt to extend the conversation with topics associated with food is likely to prove unsuccessful. This approach to pragmatics can be summed up as the view that a grammar of a language provides a characterization of meaning for each individual sentence as articulated in some semantic theory, and that pragmatic principles apply to the output of the grammar-internal characterization to yield its full import in context. This view has been justifiably influential (the Gricean view – Grice 1975, 1989). It keeps knowledge of language and general reasoning capacities quite separate – even in language use, the latter is seen as being brought into play only after the hearer has established a complete and use-independent characterization of sentence meaning. It is particularly appropriate for a Chomskyan view of linguistic knowledge as a body of knowledge which is encapsulated and independent of other cognitive capacities we humans display (see Chapter 13). Moreover, given the full array of rhetorical effects such as metaphor, irony, etc., all of which are uses of expressions in context in some sense, the proposed approach maintains a natural separation between literal uses of words, which are reflected in sentence meanings, and the various nonliteral uses to which they may be put.

2.1 Knowledge of language: sentence meanings as partial specifications of interpretation

There is, however, good reason to reject this simplistic separation of rules of semantics as part of grammar, and what I have so far called commonsense reasoning principles. The problem for this "clean" view is that we use commonsense reasoning, whatever this consists in, not merely in working out why a speaker has said something, but also in establishing what she has said in using the words chosen. Consider the conversation (5):

(5) A: Elton sang at Diana's funeral. Did you see it?
B: I spent the whole day in Kensington Gardens. The smell was amazing.

Without knowing the context of such an utterance exchange, this may seem an odd and disconnected sequence of sentences; but, in 1997, in the late summer, this might well have been an exchange between two Londoners. In the late summer of 1997, Diana Princess of Wales died in a car crash, and, after a tumultuous response from "the English people," with vast crowds gathering in mourning outside Kensington Palace in Kensington Gardens, all leaving flowers which collectively gave off what was reported to be an amazing aroma, the princess was given a state funeral. This involved a funeral process which took her coffin from Kensington Palace to Westminster Abbey, for a funeral service at which Elton John sang, a service which was simultaneously broadcast out to the many crowds outside. The questions posed by this conversation are:

- How does B understand what A has referred to by using the word *it*?
- How does talking about Kensington Gardens answer A's question?
- Why is *the whole day* taken to refer to the day of Princess Diana's funeral?
- How does A understand what the words *the smell* refer to?

The general problem we want to use this example to address is: how do language-internal principles interact with more general reasoning capacities? *It* presumably means the funeral service of Princess Diana. Since the funeral was televised, there were several means of seeing the funeral – either on television or by attending the event in person. B replies that she went to the area surrounding Kensington Palace for "the whole day." Since the funeral involved a procession from Kensington Palace to Westminster Abbey, as well as the service itself, B's reply is taken to imply that she was in the area in which the funeral took place at the time of the funeral, so her reply provides a positive answer to A's question – she was in the location at which the funeral procession took place, so in some sense at the funeral, and indeed she saw a subpart of it. B follows this reply up with the words *the smell*. She relies on A's being able to understand what these words mean by recovering information about the mass of flowers left in Kensington Gardens, and hence success in referring to the smell of these flowers. Almost none of this information is knowledge about the English language. There is nothing in the meaning of the word funeral which specifies a relation between this type of service and Kensington Gardens. There is nothing about the past tense in English which requires that the whole sequence of sentences should be taken to be about the same event.[4] And there is nothing in A's knowledge of the word *smell*, either, which specifies information about flowers left outside Kensington Palace in September 1997 – it is A's presumed knowledge of the objects so described which B relies on in choosing the words *the smell* as she does – just as it is A's presumed knowledge of the event which B relies on, in choosing to reply to A's question indirectly by referring to Kensington Gardens. Indeed, all this information is no longer generally available: the details of such events are not what get written into the history books, even if the death of the heir apparent's popular first wife gets a mention at all.

In this conversation, we see that the separation between such knowledge of language and the commonsense reasoning it drives is much more blurred than in the conversation (1). There is no sense in which B can be said to have parsed the sentence *Did you see it* and worked out the meaning A intended to convey using language-internal principles alone – only subsequently bringing into play more general commonsense reasoning principles to work out some broader message. Similarly with A in processing B's reply. The different kinds of knowledge – one language-based, the other a much more general store of knowledge – have to be combined together in understanding what the word *it*, *the whole day*, and *the smell* mean in the particular context in which they are used. However, so the argument might go, there is certainly something separate about our knowledge of language and general reasoning capacities, because B has to parse the sentence first in order to establish that the word *it* has been uttered. She has to know, for example, that *it* is a pronoun which is

the sort of word that is used to pick out some entity in the scene described in the conversation, that in the sentence A utters, the word *it* is presented as object of the verb see, and so on. Information such as this is strictly English-internal information without which B will not retrieve the meaning A intends to get across, even though none of it is itself the information retrieved. In reply, B uses the words *the whole day* relying on A's knowledge of English that an expression of this type can be taken to pick out a period of time relative to which the verb is understood. And, choosing the words *the smell*, equally, relies on A's being able to establish that *the smell* is intended as the subject being talked about, and *amazing* as the adjective predicated of it. Part of this language-specific knowledge is also the knowledge that *the* is the sort of word which leads to identifying some entity in the scenario being described; but, though words such as *it* and *the* trigger a process of identifying what is being talked about, they rely on a framework of structure constructed by parsing the sentence by language-internal rules. On evidence such as this, according to one current theory, the language system projects sentence-sized structures as sentence meanings (*logical forms*), though these are incomplete. In the case of (5) the *logical forms* corresponding to A's question and B's reply will be along the lines indicated in (6) in which some parts of the interpretation are not filled in:

(6) A: Question: The hearer saw X at time t_1.
B: The speaker spent day-Y at Kensington-Gardens at time t_2.
Smell-Z was amazing at time t_3.

X, Y, Z, Speaker, Hearer, t_1, t_2, t_3 are all parts that are missing and have to be filled in from context.[5] It is these missing parts which are transformed using general pragmatic principles of reasoning to create completed structures which more directly represent the thought that the speaker intended to convey. (Such structures are called *propositions* to distinguish them from the *sentence* that expresses them.)[6] Representations of the *propositions* expressed by A's question and B's answer might be given as follows:[7]

(7) A: Question: B saw Princess Diana's funeral on Saturday
B: B spent Saturday at Kensington Gardens. The smell of flowers outside Kensington Palace on Saturday was amazing.

The pragmatic principles which dictate how these choices are made also have a much more general role to fulfill. For, on this view, it is these very same principles that determine the broad array of metaphorical, ironic, and other effects which a sentence can convey in context. The overall picture of interpretation is that grammar-internal principles articulate both syntactic and semantic structure for sentences, a semantic structure for a sentence being an incomplete specification of how it is understood. Pragmatic theory explains how such incomplete specifications are enriched in context to yield the full communicative effect of an uttered sentence, whether metaphorical, ironical, and so on. This view is the view adopted by relevance theory (Sperber and Wilson 1995); and it is this view and the Gricean view which constitute the two major approaches to utterance interpretation.

2.2 *Knowledge of language: a set of procedures for interpreting utterances*

There is reason to think that the interaction of language-particular knowledge and general commonsense reasoning is even more pervasive than is allowed for by this modified separation of linguistic abilities and general commonsense reasoning principles. Consider a different way the conversation (5) might have gone:

(8) A: Elton John sang at Diana's funeral. Did you see it?
B: I spent the whole day in Kensington Gardens. I felt I had to. The smell was amazing. Incredibly moving.

What is it that B has conveyed in uttering the words *I had to*? How is it that he can rely on his hearer, A, to reconstruct from the word to a structure corresponding to "spend the whole day of Princess Diana's funeral in Kensington Gardens"? In this case, the speaker is giving a fragment which relies almost in its entirety on the ability of the hearer, given the context in which the string is uttered, to reconstruct some appropriate structure corresponding to the meaning of what in that context B is trying to convey. It is not that B's words themselves project a full sentence structure with lexical meanings defined independent of context, on the basis of which pragmatic principles provide some add-on means of identifying what is being talked about. Indeed, if it were words that had to be reconstructed, A would have to be reconstructing B's second utterance on the basis of the words used in B's first utterance, and that would amount to *I had to spent the whole day in Kensington Gardens* which isn't even grammatical. In case such as these, the words are mere triggers for a process of construal which has to reconstruct not only who is being talked about, but also mimic the process used in establishing the construal of what is taken as the antecedent of the ellipsis site. In the case of B's second utterance, this is achieved simply by the word *to:* It is this which leads B to understand that ellipsis site in the same way as the predicate expressed by B's first utterance, even though in this second utterance this has to be combined with the interpretation provided by the auxiliary *had*. Similarly, in processing *incredibly moving*, the hearer has to use some form of reasoning to establish what to take as the subject of the expression – is it the flowers that are so moving, or the Gardens, or the day, or the event in general? And what is the basis on which B can rely on A to build the structure into which the expression *incredibly moving* projects a predicate? There is no apparent subject to this sentence, and no verb – so where does the structure come from? Is there a rule internal to English which says that sentences with no subject and no verb are well formed? If there is, it is certainly not one which any grammar book has ever included. Evidence such as this suggests a third view. On this view, there is still separation between the intrinsic content of individual words, which is encoded (i.e., part of what an English speaker knows in virtue of knowing the language), and the process of reasoning with them, which is not encoded. Nevertheless, the process of building up the structure corresponding to some conveyed interpretation involves integrating one's knowledge of language with these general processes of reasoning at every step of the interpretation process. It is not that the rules of English syntax give rise to completed structures to which pragmatic principles fill in whatever open slots are left in the structure. Rather, we need to define a concept of structure internal to language which can be used both in building up meaning for complete sentences, and to process radically incomplete sentence "fragments" (cf. Section 4 where the question of interpreting such fragments is taken up in more detail).

3 The Process of Reasoning: How Do Hearers Ever Manage to Choose the Right Interpretation?

I have so far sketched three possible perspectives on the nature of the interaction between what we might agree was knowledge of the individual language, and more general knowledge about the individuals being described. But we have not yet begun to look at the principles which form what is arguably the center of any pragmatic theory, which explain how a hearer selects the interpretation which the speaker intended – the so-called principles of commonsense reasoning (= inference). How is it that this inferential task for the hearer manages to be successful so much of the time, given that there are many possible ways of interpreting an utterance, direct, indirect, metaphorical, ironic, etc.? What is the criterion which enables people to choose the right interpretation?

3.1 Grice's cooperative principle and the conversational maxims

According to Grice, who was the pioneer of the inferential approach to conversation (Grice 1975), there is a general assumption underpinning all utterance interpretation that the interpretation

of utterances is a collaborative enterprise guided by a "cooperative principle" in which a speaker and hearer are engaged in some shared goal. This collaborative enterprise is structured by a number of maxims, which speakers are presumed to obey, amongst which Grice isolated:

- *The maxim of quality*: do not say that for which you lack evidence; do not say what you believe to be false.
- *The maxim of relevance*: be relevant.
- *The maxim of quantity*: make your contribution as informative as is required, but not more so.
- *The maxim of manner*: be perspicuous (avoid obscurity, avoid ambiguity, be brief, be orderly).

These are not rules that dictate behavior but, rather, maxims underpinning collaborative exchange of information. Take for example marking an exam, which in England involves two examiners. I suggest to you marks for each candidate, you disagree, and we then negotiate an agreed mark. Such collaborative endeavors are said to be directed by the cooperative principle and its maxims. Of course, it is not always the case that people do tell the truth, or are relevant. And I might get fed up with agreeing exam marks and walk away. But liars and people who refuse to cooperate are in some sense the exception that proves the rule; for, in order for a lie to be successful, some presumption of the maxim of quality or its equivalent has to be in force, and people who refuse to cooperate are not engaged in any act of communication. More interestingly, the maxims are sometimes openly violated, and then they provide the trigger to a chain of reasoning which the hearer will use to reach an interpretation which the speaker intended to convey indirectly and which enables the cooperative principle to be seen to be reinstated. The conversation in (1) is an example of this. In (1), the manifestly irrelevant answer by B acts as a trigger for A to construct additional premises so that she will be led to see by indirect implication that B was trying to communicate something which is in accordance with the maxims. All such additional pieces of information, whether premise or conclusion, are said to be *conversational implicatures*. So B's answer in (1), which taken on its own is an answer that is either irrelevant or manifestly too little by way of answer, implicates that people who only know how to put a kettle on do not know how to cook, and that B does not know how to cook.

All such implicatures are derived by reasoning, and they are said to be acts of what is called nondemonstrative (i.e., nontrivial) reasoning in the face of some apparent clash with one or more of the maxims. Essential to the concept of implicature is that, unlike the intrinsic meaning of an expression, these implicatures can be "canceled" – hence their status as the result of reasoning, and not as the result of a linguistic rule. So there is nothing inconsistent with B's adding to the utterance in (1) "Though I don't mean to imply that I can't cook. I can, I'm just not very interested in cooking." If, to the contrary, some aspect of interpretation cannot be consistently denied, then by definition it was said to be part of the utterance meaning and not an implicature.[8] It is this method of retrieving interpretation through a process of reasoning in the face of an apparent violation of the maxims which lies at the heart of the Gricean account of conversation (see Neale 1996). Take for example, the much treasured compliment of being metaphorically described as a "real racehorse" in (2). This too is in blatant violation of the maxims of quality and relevance, and, in like manner, was taken by me in the situation in which it was uttered to implicate the assumptions that racehorses are extremely swift, and are exciting to watch, and that, as something described as a racehorse, I was extremely swift, and exciting to watch and hence to be with. Notice that, said with a less admiring tone of voice, what was said could equally well have been taken, much less nicely, to mean that I was highly strung, bad tempered, and easily upset. Like (1), (2) can be followed up by an explicit cancelation of at least some of these implicatures:

(9) Though I don't mean to imply that you're bad tempered or anything like that – it's just that you're exciting to be with.

Such implicatures, which in the case of (2) gave rise to its interpretation as a compliment, are said to be derived through a process of reasoning which starts from the premise that the speaker is intending to obey the general tenor of the cooperative principle, but their utterance is in transparent violation of it.[9] The cases so far considered clearly involve inferences specific to the occasion; and it isn't any rule which licenses their particular interpretation. However, there are also cases which, though they can be construed as consequences of the maxims, are so regular that it is tempting to see their interpretation as the consequence of some kind of pragmatic rule, contrary to the general Gricean spirit. These are examples such as

(10) Some people there were miserable.

(10) would normally be taken to imply "Not everyone there was miserable," but it is cancelable as in (11):

(11) Some people there were miserable. Indeed everyone was, though some were showing it less than others.

Since it is cancelable, according to the criterion defining an implicature, it is not an encoded principle of the grammar. Grice labeled implicatures such as these *generalized conversational implicatures*, but others since then have given them a rule-based characterization, suggesting that the concept of a grammar of a language might be extended by a pragmatic component which contains a set of default rules (see Gazdar 1979; Levinson 1983, 1987, 1996; and Lascarides and Asher 1993; Asher and Lascarides 2003, for a concept of the *default inference rule*).

One primary difficulty with these maxims of Grice's is that they are often not clear, and any single implicature can be reasoned to be a consequence of a number of maxims. Is, for example, B's answer in (1) to be construed as a violation of the principle of relevance, or of quantity? Has B said too much, or not enough, or merely something irrelevant? Any one of these could be taken as triggers to the chain of reasoning that leads to the intended interpretation that B cannot cook. Then there is the problem of what it means for some utterance to be relevant, a question to which Grice provides no answer. The result is that, though suggestive, the content of the maxims remains extremely vague. A further problem is: What should the maxims be taken to explain? Grice articulated the maxims as a means of simplifying the overall account of the relation between the use of language in logical arguments and the conversational use of language, arguing that recognition of maxims of conversational behavior could be used to explain a well-known problem – the apparent mismatch between how words such as *and*, *or*, *not*, *if–then* are used in logic and their use in ordinary language. In logic, which is the formal study of reasoning, the focus is on arguments that are valid in virtue of structure, for example that displayed in the English sequence:

(12) If Bill is married to Mary and Mary is a Professor of English, then Bill is married to a professor of English. Bill is married to Mary, but Mary isn't a professor of English. Therefore Bill isn't married to a professor of English.

Displaying the validity of argumentation in a sequence such as (12) involves defining *and* in such cases as combining two statements to yield a further statement of the form *P and Q* which is true if and only if the first statement and the second statement Q are both true. *And*, that is to say, has a purely cumulative effect in the steps of reasoning involved: nothing turns on the order in which the two statements are written. This use of the word *and* is unlike its characteristic use in conversational sequences, where it is often associated with sequencing in time. (13) for example does not merely imply that Bill was sick at some time in the past and that Bill went to bed some time in the past – it implies that he went to bed after he was sick, a sequence of events reversed in (14):

(13) Bill was sick and he went to bed.

(14) Bill went to bed and he was sick.

Examples such as these were used in the 1950s to 1970s (Strawson 1952; Cohen 1971; Walker 1975) to demonstrate the difference in content between natural language expressions and elements of logical languages, with the logical concepts being defined in terms of conditions necessary for the truth of a given element (e.g., *and* as in (12)). Grice however argued (1975) that the difference between *and* as in (12) and as in (13)–(14) was merely a difference in the implicatures that can arise as a result of the cooperative enterprise underlying the process of communication.

This insight of Grice's was taken as a major advance, because it enabled natural language content to be defined in terms familiar from logic despite the superficial divergence of natural language interpretation from that defined in logic, with the result that natural language interpretation was thereby taken to be better understood, divided up into two distinct phenomena, the concept of meaning for sentences, and the concept of meaning for utterances. Following this methodology, the meaning of sentences was defined in terms of truth relations, the meaning of *It's raining* being said to be given by the set of conditions which have to hold in order for the sentence to be true. This gave rise (Gazdar 1979) to the slogan:

Pragmatics = meaning – truth conditions

Pragmatics was seen as the heterogeneous remainder left over once the account of descriptive content in the form of truth conditions is articulated as the basis of semantics – this left-over being explained in terms of maxims of behavior which provide the trigger for the "common-sense reasoning" process that adds to (or replaces) the stricter descriptive content of an uttered sentence. Notice how the slogan itself leaves open the question of whether these maxims should be characterized as constraining an inferential task (as Grice 1975 advocated), or as default rules as some of his followers have assumed (see Brown and Levinson 1987; Levinson 2000). As indicated earlier, the Gricean program is attractive to linguists as it buttresses the view that there is a clear separation between grammar-internal processes, which characterize sentence structures, and arguably also a specification of their meanings, and the interpretation of utterances (see Brown and Levinson 1987; Atlas 1989). It is attractive to semanticists also for the same reason: It allows concepts of truth-conditional semantics familiar from the study of logic to be extended to sentences, defining the domain of pragmatics as a form of explanation that takes as input such specification of sentence meanings (Kamp 1978).

There are, however, serious difficulties with this view. The first is most obviously displayed by pronouns and other anaphoric expressions.[10] In order to establish the truth-conditional content expressed by a sentence containing a pronoun, some choice as to how the pronoun is to be interpreted has to be assumed; and these choices are not given as part of the grammar of a language in any sense – they depend on the interpretation of the sentence as understood in a context. The conditions under which (15) is true are, for example, quite different depending on whether the pronoun she is taken to refer to Princess Diana's body, or to each individual woman being talked about:

(15) [*uttered in follow-up to (5)*]: Every woman cried as she went past the gates.

If *she* is construed as picking out Princess Diana's body, it picks out a fixed object: if she is construed as ranging over the same set of individuals as is picked out by *every woman*, it picks out a set of individuals, each one in turn. The two circumstances described are very different. Notice, too, the way the interpretation of the pronoun varies may be sensitive to details of the particular event described. The funeral procession that morning in 1997 started from Kensington Palace, where the princess had lived, and that morning no one came through the gates until the beginning of the procession. If the speaker and hearer can both presume on this sort of knowledge, then in (16), *she* will be construed as the princess' body. However, in (17) *she* will be construed as each of the women being talked about, because it was bystanders who put flowers beside the gate (hardly the dead princess!):

(16) Every woman cried as she came through the gates.

(17) Every woman cried as she put down her flowers.

The problem that truth-conditional content depends on specifics of individual contexts is by no means restricted to anaphoric expressions such as pronouns. The women being talked about in an utterance of (17) as follow-up to (5) – who are assumed to be picked out by the speaker's utterance of the expression *every woman* – would be tightly restricted to the women described outside Kensington Palace as seen by B. It could not be every woman who went that day to put flowers down on some other grave elsewhere in the country, nor every woman at home watching the event on television. Even the predicate *put down* does not describe every event of putting down flowers that took place that day. Many people at the funeral event may have put down their flowers as they adjusted buttons on their coat, wiped a child's nose, etc. etc. – let alone women right around the world. But *put down* here refers only to the act of putting down flowers outside the gates of the Palace. Indeed the interpretation of *any* of the words may in part be due to the context in which the sentence is understood. From this we are driven to conclude that grammar-internal principles do not determine full specifications of truth-conditional content but much less complete specification; for "commonsense reasoning" principles are also dictating what is expressed by a sentence as uttered in a context. Grice himself did not see the construal of pronouns and other anaphoric expressions as a major problem, and, within the Gricean concept of utterance interpretation, this was not seen as controlled by the maxims (see Carston 1998 for discussion). However, as we shall see, the very same criterion that determines how indirect implicature effects are recovered by the hearer also determines how all such context-dependent aspects of interpretation are chosen. Finally, as Grice himself pointed out (1975, 1989), there are some aspects of meaning projected by words that are signally left out of any program that defines the meaning of natural language expressions in terms of truth-conditions. These are aspects of meaning which do not have anything to do with properties of external objects that the word can be used to describe, but, rather, have to do with the mode of reasoning about such objects that the word triggers. Take the word *but*, used in the previous sentence. The word *but* is used to indicate some form of contrast, but this is not a contrast of content intrinsic to the entities described. We know that this is so, because of examples such as (18) and (19) where explicit identity of the predicate is asserted in the two clauses joined by *but*:

(18) John is applying for the Liverpool job, but so am I; and I have more publications than he does.

(19) John got 70 percent but so did his brother, so neither of them could boast about having done better.

The truth-conditions contributed by the use of *but* cannot be distinguished from that of *and:* a statement formed by joining together two statements by *but* is true if and only if the two statements are true, exactly as in the case of *and*. Any characterization of word meaning merely in these terms will miss the idiosyncratic contrastive flavor intrinsic to the meaning of *but* altogether. Yet the phenomenon is not a conversational implicature either, as it is an invariant aspect of sentences conjoined by *but* that some form of contrast is intended to be recovered. The Gricean program has to allow for special stipulations for aspects of meaning such as these which fall outside both the truth-conditional program and the implicature form of explanation, as Grice himself pointed out (1975). He called these phenomena "conventional implicatures" to indicate that they were not regular implications of descriptive content but nonetheless part of the conventional meaning of the word in question; but this term was little more than a classificatory label.

3.2 Relevance theory

These various challenges to the Gricean program were taken up in Relevance Theory (Sperber and Wilson 1995). This theory claims to characterize pragmatic phenomena in terms of a single

cognitive concept, that of relevance, replacing the social underpinnings of Grice's cooperative principle.

3.2.1 The principle of relevance According to Sperber and Wilson, there is one overriding constraint on human cognition, the principle of relevance. All signals are said to be processed relative to an ever-evolving background context with the aim of retrieving more information from the signal than it itself presents; and optimal relevance is getting the right balance between size and type of context and amount of information derived for the task at hand. So, in our Princess Diana example, we might say that the interpretation of the word *it* in (5) picks out the funeral because this has just been mentioned in the context and so costs no effort to recover, and that referring to Kensington Gardens and the smell of the flowers gives rise to a whole chain of impressions and so is highly informative in talking about the funeral. Hence the relevance of each succeeding sentence in the exchange in (5). This trade-off between cognitive effort and cognitive effect is at the heart of the concept of relevance itself. Humans are said to always subconsciously balance the amount of effort a task should have against the benefits to be gained from it – interpreting incoming signals to get as much information from them as is possible relative to putting in the minimum effort necessary to achieve that effect. This balancing of effort and inferential effect is the constraint of maximizing relevance. The more information some stimulus yields, the more relevant it is said to become, but the more effort the interpretation of that stimulus requires, the less relevant it will become. And to be minimally relevant a stimulus must lead to at least one nontrivial inference being derived.[11]

To take in more detail a nonlinguistic example first, imagine yourself trying to write a letter in reply to some job advertisement which needs to be got off today if you are to have any hope of being treated as a serious candidate. Imagine also that it is raining. To you, the information that it is raining is not relevant because you are trying to work out how best to present yourself and your achievements. This can be characterized through considerations of amount of effort required for the inferential effects to be achieved. What you are currently focusing on – the context against the background of which you are constructing your application – are premises which concern this particular act of writing: "I must remember to mention my degree results," "I must remember not to use too many adjectives," "If I use the word *impressive* too often, they will think I am boasting." And so on. What the weather is like does not impinge on this activity, at least not just at the moment. The effort of retrieving and manipulating information about umbrellas, whether to go by train, car, or bus, is not warranted relative to your current worries, for none of these premises will combine with premises about how best to communicate what an impressive individual you are. When, however, it comes time to go out of the house, then there will be decisions to be made, and these involve reasoning with premises about the weather and the nature of the journey to be made; and the information that it is raining will combine with these to yield appropriate inferential benefits (for example, if it were me I would be thinking things like "It's better to go by train and read, since the rain is likely to mean that going by car will be slow"). The cognitive effort of drawing such inferences at this later point in time is suitably rewarded, in the sense that noticing that it's raining combines with other things that are then on your mind anyway. According to this approach to understanding, the interpretation of a stimulus is defined as the manipulation of additional information relative to which a nontrivial set of inferential effects are achieved. Interpretation of signals of all sorts on this view invariably takes place relative to a context; because context is defined as the premises selected (the extra information) which ensure the relevance of a signal. There is no concept of a null context. This example only involved one person, working away on their own, not being distracted by the rain. However, interpretation of an act of communication involves two agents – the speaker and the hearer. The constraint of balancing cognitive effect (the drawing of nontrivial inferences) with cognitive effort will also apply to what the hearer does, but here the task of interpretation is more specific because the hearer has to try and recover what the speaker intended to convey.

There are two aspects to the task of utterance interpretation on the Relevance Theory view. First, there is the decoding of information intrinsically associated with an uttered expression – i.e., working out what words have been said and the information that they by definition carry. Second, there is the need to make choices which enrich that encoded information to establish what the speaker had intended to convey using those words. To succeed in the first task, one has to know what the words of the language mean – what information they encode. This is the starting point from which the proposition expressed has to be recovered. To succeed in the second task means establishing (a) some proposition corresponding to the intended interpretation of the utterance, (b) additional propositions which establish the required inferential effect. These propositions are, however, not just those which happen to be maximally relevant to the hearer: they must at least include those which the speaker could have intended. And this is where a somewhat different principle of relevance – the communicative principle of relevance – comes in. A presumption of optimal relevance, according to Sperber and Wilson (1995), determines how the hearer succeeds in arriving not merely at some most informative interpretation, but at the interpretation which the speaker intended. The context against which these decisions are made is said to be the set of representations retrievable with least effort that establish requisite inferential effects (this is what corresponds to "optimal" relevance).[12] This set may be taken as containing just the immediately previously constructed proposition: and, indeed, direct answers to a question do combine with such a context, viz. the question itself, for which they provide an answer. However, the context selected could also be some extension of the minimal context, as long as the extra effort required is offset by additional inferential effects; and this is what is triggered in indirect replies to questions. For example, in interpreting the uttered sentence *I went to Kensington Gardens* in response to *Did you see the funeral*? in (5), the proposition taken to be expressed will be "B went to Kensington Gardens." This indirect response, despite the increased effort required to process it, would have the advantage of triggering an extension of the context to include a premise such as "Kensington Gardens was the starting point for the procession preceding the funeral" from which A will deduce that "B saw the funeral procession by going to it" and probably also that she saw the service as well on the huge screens that were made available so that everyone outside the abbey could watch it. The benefit of the indirect answer here, according to relevance theory, is that for a minimal increase in effort of processing the given input, the hearer A is recovering extra information which she would not otherwise have got – here, the much richer stock of information that B did not see the funeral by watching it on tv, that B was part of a historic event, that B has much more information about some aspects of the event than she does, that B will not have seen other parts of the event, etc. Hence B's answer in (5), though indirect, and causing A more effort in parsing and constructing an interpretation, is optimal in guiding the hearer to the requisite range of inferential effects. The very indirectness of B's answer indeed is intentional, allowing an open-endedness in the interpretation since the choice of context is not fully determined.

3.2.2 Relevance and the recovery of what is "said" One immediate advantage of the relevance-theoretic approach over the Gricean one is that the explanation of how implicatures intended by the speaker are worked out applies equally well in explaining how the proposition the speaker has expressed is arrived at. So, for example, (16)–(17) uttered in the context of (5) are predicted to give rise to different interpretations. Choice of *she* as the variable bound by *every woman* in processing (16) would not be possible because it would combine with the information that no woman went through the gates other than the dead woman in her coffin, giving rise to immediate inconsistency, and no further nontrivial inferences.[13] All such interpretations are therefore ruled out as not relevant. Hence *she* can only be used in that context to refer to the individual inside the coffin. Similarly, though to reverse effect in (17), *she* cannot be construed as Princess Diana since neither dead people nor their coffins can lay flowers. In both cases, the only available choice of representation is the one that is selected – a representation which meets the criterion of giving rise to a consistent set of inferential effects without undue cognitive effort. Exactly the

same constraint dictates the construal of *put down* as "put down outside the gates of Kensington Palace," though this time for a different reason. It is perfectly possible that an individual can cry as she puts down her flowers to wipe a child's nose or do up her coat, but in this case, the type of premise required is one that combines with an assertion about everyone there. Information about women except in relation to events concerning the activities involved in the mourning at Kensington Palace has not been made salient and so is not easily recoverable. In contrast, the selection of *put down* as "put down beside the Kensington Palace Gates" in that same context naturally triggers such easily available premises as: Crying is an explicit gesture of mourning. Putting down flowers by Kensington Palace was an explicit act of mourning. Such a choice of context would give rise to the inferential effect Every woman who put down flowers by Kensington Palace did at least two things as an explicit gesture of mourning. Since such a set of premises is easily recoverable, the interpretation of *put down* as "put down by Kensington Palace" is the concept the speaker intended to convey.[14] Similar relevance considerations dictate that *every woman* picks out women standing by as mourners rather than, for example, women watching the same event at home on tv.[15]

So the principle of balancing cognitive effort and inferential effect can be seen to underpin both the deduction of so-called implicatures and the fixing of context-dependent aspects of the proposition expressed. It has the advantage also of not requiring the explanation of additional pragmatic effects to be triggered only in the event of apparent violation of a conversational maxim, as does the Gricean account. Indeed, it purports to explain why deduction of additional information is an unvarying consequence of interpreting an utterance, and not merely a feature of exceptional apparently anomalous conversational exchanges. Moreover, it provides a natural distinction between implications which the hearer believes the speaker intended to convey (= *implicatures*), and those which she recovers from the utterance despite knowing that the speaker could not have intended to convey them (= *contextual implications*) (Sperber and Wilson 1995; Carston 1988, 2002). The implications the hearer believes the speaker intended to convey are those dictated by the criterion of optimal relevance (with minimal cognitive effort as the overriding factor). Those which the speaker need not have intended are the result of the less restrictive criterion of maximizing relevance. These will often add very considerably to the relevance of the overall utterance for the hearer, though they cannot be taken to be part of what the speaker has intended to convey. The fact that B went to the funeral, for example, will provide A with information about B, that B got completely caught up in the fervor that swept the country that weekend, that therefore she is probably a traditionalist, and so on – none of which B would have explicitly intended to convey in her reply. Such implications, by the way, would not be characterized by a Gricean system as part of the utterance interpretation process at all.

3.2.3 Relevance and speech acts We have so far assumed that information retrieved from an utterance is solely about the object referred to by a speaker. However, humans fluently reason at both the level of what is communicated, and at the level of how and why something is communicated. So A in our first conversation (1) will not only retrieve the information that B knows how to put a kettle on, but also that B wants A to believe that she only knows how to put a kettle on. Such higher-level information is partly encoded – the differences between assertions, imperatives, and questions, in particular, rely on our ability to retrieve such higher-level implications. An assertion implies that the speaker, if taken sincerely, believes the proposition his utterance expresses is true. An imperative is a request by the speaker that the hearer make some proposition true. A yes–no question is a request to the hearer to indicate to the speaker whether some proposition is believed by the hearer to be true. These higher level *explicatures* (as they are labeled in relevance theory, see Carston 1988) have been studied as part of semantics under speech act theory (initiated by the philosopher J. L. Austin in Austin 1962) and, following him, Searle and Bach and Harnish (Searle 1969; Bach and Harnish 1979), and are an important mechanism for achieving coordination between conversational partners. In this earlier theory (which predated Gricean

pragmatics and relevance theory), language use was described in terms of a range of speech acts in terms of which the meaning of natural language expressions was explained.[16] There are many cases where we do much more with words than merely describing objects and activities around us. So, for example, when a minister holds a baby over a font, sprinkles a few drops onto his head, and says the appropriate words from the baptism service, he is not merely saying these words, he is carrying out the act of baptizing. And when I say "I promise you that I will send you a letter tomorrow," I am not merely saying those words – I am carrying out an act of promising.

On the speech act view of language, language can best be understood in terms of acts such as these which speakers carry out in using language. The observation by speech act theorists that there is clearly more to language than just describing things is quite uncontentious – no one working in pragmatics doubts this. Nonetheless, in relevance theory, where the type of implications that can be drawn is quite unrestricted, there is no need of any special discrete categories for such different kinds of act. All these implications would fall under the category of *explicature*, and as part of the proposition expressed would come within the general umbrella of information retrievable from an utterance, for which the hearer's task is to recover those implications that the speaker intended to convey. Like all other implications, whether or not they are retrieved depends on their relevance to the participants. It may be relevant to construct the explicature that the speaker believes that P, for some arbitrary P, though in ironical utterances this is contra-indicated. Questions normally require an answer, but so-called rhetorical questions signally do not. And, equally, it may be relevant that someone we speak to has specifically promised that P. Even in the highly conventionalized case of baptism, the implication that a child has been baptized through some specific act in a religious service is potentially relevant in just the same way as every other act of communication – it needs no special pragmatic category to explain its communicative effect – merely an explanation of the role of the priest within a given religious ritual and the significance of baptism within a set of religious practices. I shall not have anything more to say about such "speech acts" in the remainder of this chapter, but they nonetheless have an integral role in establishing the relevance of an utterance (see Wilson and Sperber 1988; Carston 2002).

3.2.4 Procedural aspects of interpretation The assumption that interpretation involves constructing both some distinguished proposition and some context set of propositions to combine with it provides a natural basis for explaining the conventional implicature phenomena problematic for Gricean approaches to meaning. Given the two-fold nature of the inferential task triggered by natural language input, it is entirely natural that the content of some words might be directed more towards constraining the context set of premises to be selected, rather than in establishing the proposition expressed (see Blakemore 1987, 1992). In this light, we can view connectives such as *but* as constraining the context relative to which the sentences it connects are to be construed, establishing both one form of context for the first conjunct and a guaranteed modification of it in adding the second conjunct. Seen in these terms, *but* has to be defined as a procedure for context construction, imposing a choice of context for the first conjunct which must lead to a conclusion, relative to which the context for the second conjunct (which will automatically contain the first conjunct) must yield a contradiction (Blakemore 1989).

Since the important work of Blakemore (1989) it has become increasingly evident that the concept of procedures for interpretation has much more general application than merely the set of clause-joining connectives (see Carston 2002; Wilson and Carston 2007; Kempson, Meyer-Viol, and Gabbay 2001; Cann, Kempson, and Martin 2005). Consider first pronouns. Pronouns do not fix the way the hearer understands them; they merely guide the utterance process. This guidance takes the form of a constraint. Consider (20)–(21):

(20) The queen frightened her.

(21) The queen frightened herself.

(21) must mean that the queen frightened herself, rather than anybody else. (20) however cannot mean this – it must mean that it is someone other than herself that the queen frightened. More formally the pronoun provides a placeholding device for which the hearer has to choose some representation, relative to the constraint that the individual represented be female *plus* the constraint that the representation the hearer selects must not be one that is already set up within the propositional structure being constructed. It is for this reason that (20) and (21) cannot be taken to mean the same. This "locality" restriction is generally seen as a syntactic restriction defined over syntactic structure, separating some aspects of pronoun interpretation off from their use in discourse as grammar-internal and subject to syntactic explication (see Chapter 13 and Chomsky 1981); but, looked at from a more procedural perspective, we can see the pronoun as encoding a constraint on the building up of a propositional structure (see Kempson 1988a, 1988b; Kempson, Meyer-Viol, and Gabbay 2001; Cann, Kempson, and Martin 2005) – a procedure which guides the hearer but leaves open what it is that the pronoun actually picks out on any occasion of use. But there is an important consequence to this shift. If the intrinsic lexically specified meaning of a pronoun is to be defined as a set of procedures for interpretation, it is essential that interpretation be defined in terms of structured representations of content, as locality conditions have to be defined over a level of representation, and this can be construed as an encoded constraint on pragmatic interpretation *only* if the process of interpretation is also taken to be defined in terms of structure.

There is reason, however, to go further than this. Remember first how the interpretations of *the smell*, and *the whole day* in (5), *woman* and *put down* in (17) also depended on context, displaying a similar gap between the characterization of the word independent of context and its particular interpretation in a given utterance. The very generality of this phenomenon of context-dependence suggests that all specifications of words should be viewed, in some sense, as constraints, underspecifying interpretation, on the basis of which a hearer builds some propositional structure.[17]

It turns out that such a shift has a big advantage, as it enables metaphorical construal to be brought into the general remit of semantic theory as merely an extreme illustration of how words can be used to construct entirely occasion-specific concepts (*ad-hoc concepts*: Carston 2002). There is an important background debate behind this suggestion. There has been disagreement within pragmatics over whether there should be any independent maxim of quality constraining people to tell the truth (Wilson 1995). The maxim of quality contains the heart of the social principle of cooperation intrinsic to Grice's theory of conversation. On the Gricean account, remember, the interpretation of metaphor involves transparent violation of the maxim of quality with the false proposition expressed by the sentence as its literal meaning having to be replaced by some quite different proposition (see Sperber and Wilson 1981 for a critical evaluation of the Gricean account). The Sperber and Wilson account, however is *not* a social theory of communication; and, as currently now being developed by Wilson and Carston (2005), metaphorical interpretations are merely an illustration of the constructive process that the words trigger, allowing the relation between the thought which the hearer recovers from the sentence as uttered in context and the words used to convey it to be really quite indirect. On this view interpretation of a sentence such as (22) involves constructing a concept on the basis of the presentation of the word *cornered* in such a way as to yield extra information about the entities picked out by the pronouns in that context, in ways which make the whole utterance relevant to the hearer.

(22) She cornered it.

So *cornered* has to be understood as a relation between individuals, something someone can do to something else of the sort we describe by the verb to *corner*. The construction of some occasion-particular concept is direct and not via a process of rejecting a literal interpretation. No maxim of quality needs to be invoked in this account of metaphor, but, equally, no indirect concept of resemblance is invoked either: It is, rather, that all that a word provides is a set of procedures, or some template relative to which a hearer establishes a relevant

interpretation, constructing new concepts online from what the word presents. In this case, supposing the pronoun *it* refers to her car, one will interpret the verb as conveying a relation between this woman, whoever she is, and her car such that she gets it into or round an angle in some road or track. Exactly the same procedure in principle would apply to explaining the construal of more obvious metaphorical examples such as (2) where the adjective *real* has to be taken as triggering some concept of *racehorse'* and the attributes that this type of "racehorse" display.

It may seem that such quasi-metaphorical or metaphorical uses of language are peripheral, and so not strong evidence that all linguistic meaning should be in these more indirect terms. However, this is far from the case. Every time we process a tensed sentence, we find that we construct an occasion-specific, in this sense ad hoc concept. Consider what we convey when we say:

(23) I watched the birds in Hyde Park yesterday all afternoon with my grandson.

In processing such an utterance in an ongoing discourse, the hearer is first presented with words that induce the construction first of the concept "watch" as applied to some group of birds, a relation which immediately has to be refined to include the information that this event was located in Hyde Park (in central London), an event which is then described as taking place the day immediately before the day in which the sentence was uttered, and then, extending the description of the event, the hearer comes to understand that this birdwatching event extended throughout the afternoon of that day, with the final piece of information that is added being that the speaker was with her grandson. This last adjunct adds its own flavor to the understanding of the concept of birdwatching originally constructed from the words *watched the birds* depending on the hearer's knowledge of the age of the grandson (birdwatching with a two-year-old is a very different event from birdwatching with a keen birdspotting teenager!). As this example shows, coming to understand what type of event located in space-time is described depends on the incremental processing of each adjunct phrase, with each having to be interpreted as dependent on context for the particular fixing of what it is that it contributes to that description. So the phenomenon of building up arbitrarily complex situation-specific concepts from the words is essential to understanding what the speaker is communicating. All that the words provide are the procedures which enable such build up of interpretation to take place.[18]

4 Grammar as Defining Procedures for Proposition Construction

Having gone this far with the concept of procedures for proposition construction, we can go even further. Syntax can also be seen as the articulation of mechanisms for building up representations of content (Kempson, Meyer-Viol, and Gabbay 2001; Cann, Kempson, and Martin 2005). To make this move, we need to be just a bit more specific about what we take such structured representations of content to be. Notice in Figure 20.1 we can represent the content recovered from some uttered sentence in a tree-structure format which indicates the content of the proposition (decorating the topnode of the tree), but also its parts and how they have combined (each argument by convention put on a left branch and the predicate nodes corresponding to the verb-phrase meaning, and that of the verb itself on the right branch):

What we then need to define is a process of how such a tree can be progressively built up incorporating constraints on interpretation that the words provide. To define this, we need a few basic strategies for building trees, and then to define word-contents in such a way that they can be seen to interact with such strategies in a natural way. The general strategies we need are ones which initiate the building process from some presumed starting point. We take the point of departure to be a one-node tree, which states the goal of building up a proposition, which we might

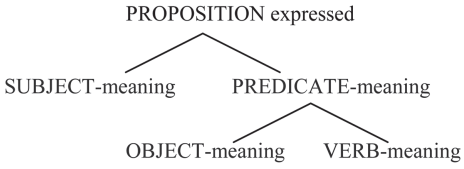


Figure 20.1 Proposition-internal structure.

write as *?PROPOSITION*, or more simply *?t*, where *?t* Indicates that the node is required to be a proposition (*t* is for a truth-value denoting type, hence a proposition; and the '*?*' indicates that this is a requirement that has yet to be met). To this we add a small number of opening moves. One strategy is to allow that such a tree can be expanded with two subgoals, the goal to build up a subject *?SUBJECT* and the goal to build up a predicate attributed to that subject *?PREDICATE*. We write the former as *?e*, where e is the category of individual-denoting entities and, as before, *?* indicates that the provision of a formula of the right type has yet to be provided. The predicate requirement is then a requirement for some category that will apply to the subject category to yield a proposition category; and this is the requirement $?e \rightarrow t$. (See Heim and Kratzer 1998 for an introduction to the concept of semantic types.) Such an expansion will give us as a possible opening sequence the partial tree in Figure 20.2:

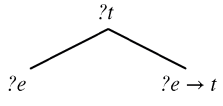


Figure 20.2 Subject-predicate structure.

Turning now to the kind of information the words provide, let's look at the building blocks for a sentence made up of an intransitive verb and its subject, such as in (24):

(24) John sneezed.

The words can themselves be defined as projecting partial structures, which can be used to enrich such an initial opening sequence. For example, we might analyze an intransitive verb such as sneeze (see also, *faint*, *swim*, etc), as projecting an entire skeletal proposition. That is, we can see the verb *sneezed* as projecting the structure in Figure 20.3. This is no more than the assumption that it is verbs that create primary structure of the propositions which our sentences can be used to express.

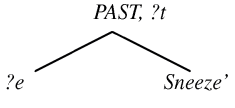


Figure 20.3 Lexical information provided by *sneeze*.

On the other hand, names in natural language can be defined as inducing the construction of a representation uniquely picking out some individual, with the word *John* associated with a representation uniquely picking out some individual for whom there is an established practice of picking him out with that name. All such information, both for nouns and for verbs, would be stored in the lexicon.[19]

To get these two pieces of information to combine together in a procedure reflecting real-time processing, we now assume the selection of the subject-predicate strategy that introduces a partial structure requiring a subject and requiring a predicate, and we use this as input to parsing the word John. This would give us the tree in Figure 20.4.

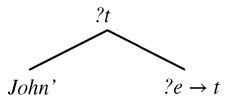


Figure 20.4 Decorating the subject node.

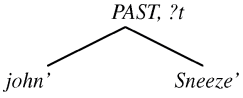


Figure 20.5 Processing subject and verb.

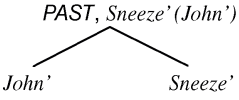


Figure 20.6 Achieving a propositional formula.

This tree can then provide input to the parse of *sneeze*; and because the frame which the word sneeze provides is directly compatible with this, it can be used to extend it creating the update tree in Figure 20.5.

Now, because the predicate *Sneeze'* is of a type that combines with an individual type of expression to yield a proposition, the two formulae decorating the daughter nodes can be combined (by so-called functional application, see Heim and Kratzer 1998) to yield the tree in Figure 20.6.

With just incorporation of tense information (which we gloss over here), we have achieved the shift into seeing syntax and lexical information in combination as projecting propositional structure. Transitive verbs, such as *hit*, are just a slightly more complicated variant of this pattern, triggering a process for constructing a structure containing two positions to be filled, one to be filled by an object and the other by a subject; but the general stance is preserved (Kempson, Meyer-Viol, and Gabbay 2001; Cann, Kempson, and Martin 2005).

Anaphoric expressions, *he*, *him*, *himself*, *the man* and their construal, fit into this perspective seamlessly, as we have already seen that these can be seen as procedures for building up such structures relative to constraints that determine how the various arguments are to be identified (in the case of reflexives as forcing the identity of two arguments within a propositional structure, in the case of pronouns precluding any such co-argument identity). Such an analysis of reflexives and anaphora more generally is entirely uncontroversial: all that is unorthodox is the assumption that syntax itself should be perceived in similar procedural terms, defined directly in tandem with language processing as constraints on the incremental unfolding of predicate-argument structure following the dynamics of left-right processing.

More surprisingly perhaps, this interpretation-building perspective can even be applied to structures that have been identified standardly as requiring an independent level of syntax in virtue of dependencies which are not local in this sense. In particular it can be applied to the case taken to be the heart of natural language syntax, the phenomenon displayed in English by the *topicalization* structures (25)–(26):

(25) John, Mary upset.

(26) Which sonata do you think John is expecting to play at the concert?

What we need for this type of case is the addition of one more basic strategy. This is a general strategy for the early stage in the building up of the emergent prepositional tree, specifically for the first word or constituent in a clause, in which the structural relation may not be fully determined, as indeed in (25)–(26). At the outset of parsing this type of structure, a structure-building device is needed that gives only the information that, within this newly emergent propositional structure, a new node is to be constructed with the information that, the node in question will at some later point have a position within that tree, a position which is not yet determined. Slightly more formally, within a proposition-requiring tree, there is a licence to construct a tree-relation that is unfixed within that proposition-requiring structure (represented by a dotted line in the emergent partial tree in Figure 20.7).

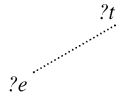


Figure 20.7 Constructing an unfixed node.

With just these assumptions, we can now show how topicalization structures can be progressively built up. We do this with a parse of (27), as this will show at the same time how the strategy interacts with the task of building up interpretation for a pronoun, the central case of an anaphoric device:

(27) John, she admired.

All we have to assume is that, unlike the parsing of (24), the unfixed node which the initial expression can be used to decorate can be resolved *anywhere* within some subsequently emerging structure; and, in particular, it is not immediately resolved as the subject (indeed, it can't be, in the construal of (27), as there is no immediately following verb). So the first word *John* is taken to decorate an unfixed node exactly as in the parse of (24) (see Figure 20.4). The pronoun *she* on the other hand does provide information, i.e., the decoration it provides must be identified as subject. Moreover, in the context in which the interlocutors, let us suppose are talking about Sue, the term *Sue'* can be selected as the interpretation of the subject (this of course depends on the context). Furthermore, subjects, by definition, demand a predicate, so the hearer is able to establish the structure as in Figure 20.8 (indeed there is in this tree-growth perspective independent reason why such a nonadverbial NP immediately before the verb in English must be construed as subject, but we leave that here: see Cann, Kempson, and Martin 2005; Cann, Kempson, and Gregoromichelaki 2009).

At this second stage, what the hearer has established is not merely that *she* is the subject picking out Sue but also that there must be some property attributable to Sue which will combine with that term to form some requisite propositional formula (hence both the node with a $?e \rightarrow t$ in Figure 20.8, a predicate-requiring node, and the requirement *?t* decorating the top node).

With the parsing of the subsequent word *admired*, more information can be established, as verbs, as we have already seen, induce propositional templates specifying how many arguments their associated predicate has to combine with. Accordingly, the structure in Figure 20.8 can be enriched to yield Figure 20.9.

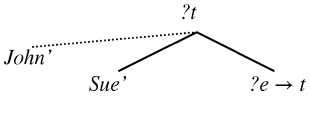


Figure 20.8 Parsing *John Sue* in (28).

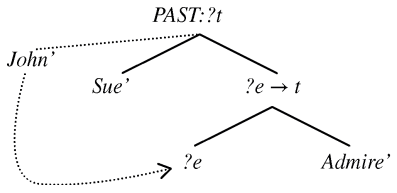


Figure 20.9 Parsing *John, she admired.*

Again, there is a predicate node not yet established, but, this time, there is sufficient information to enable the object argument of the predicate *Admire'* to be identified. This is achieved by unifying the unfixed node with what had been previously been an open as yet unfilled argument annotation, and, with this move, the proposition taken to be expressed by (28) can be established as 'Sue admired John' (Figure 20.10).

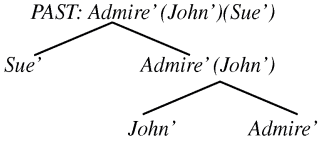


Figure 20.10 Completing a construal of *John, she admired.*

Notice what form the information presented by the words has to take in this step-by-step modeling of the interpretation process. They have to be seen as procedures rather than simply as providing conceptual content. What the mechanisms provide, whether by general strategy or by lexically provided definitions, are the means to progressively update an initially very partial tree. Both lexical and syntactic information, that is, induce the construction of a proposition relative to information independently available in context.

Of course, there is a very great deal more to be said to cover auxiliary verbs, adverbs, adjectives, etc., before even a basic case is set out, and this difference between frameworks cannot be resolved by one somewhat oversimplified derivation. Structures like these have been the focus of much study as part of core syntax, where the metaphor, over many years, has been that the representation of *John* has moved from some original position as an argument subcategorized by the verb hence, at that level, adjacent to it, to the position in the front of the string, leaving what is called a "gap" (see Chapter 13 on syntax). Indeed, there is one standard argument that syntax has to be defined independently of interpretation, but the reply is as easy as the objection. It is well known that these "movement" processes are subject to restrictions as to what structures they can be "extracted out of," and this is said to constitute a purely syntactic restriction not explainable in terms solely of semantic content. However, all we need to express such restrictions is the concept of structured representations so that notions of locality can be defined over suitable-sized structures, and such a restriction is just as expressible in a framework in which what are built up for interpretation are structured representations of content (see Cann, Kempson, and Martin 2005).[20]

Setting this problem aside here, there is a sense in which a movement approach and the approach with more of a parsing flavor constitute two sides of the same coin: Viewing the internal mechanisms of language as themselves constituting the core of the parsing process is little more than the inverse of the posited movement process which takes as input some identified object–verb relation. However, it does give a different basis for explanations. On this parsing-oriented view, concepts of underspecification are central: In any parse activity, at the outset, there is very little information; and it is from that initial point of departure that there is gradual build up of information as the mechanisms that each word makes available are made use of until such point at

which the information provided by this initial expression can be slotted in, as is displayed by the transition across partial trees set out in Figures 20.1–10. This has the advantage of corresponding very directly to the dynamics of utterance interpretation, a task to which the grammar has, in some sense, to be applied.

4.1 Syntax mechanisms and ellipsis construal

The significance of the claim being offered is not so much that parsing involves progressive buildup of structure relative to how the words are understood, for this is uncontroversial, but that this should be taken as the basis for explaining syntactic distributions. Confirmation of this stance comes from ellipsis construal, which displays what is otherwise a puzzling mixture of structural, semantic, and pragmatic properties.[21]

Elliptical constructions are those where something is in some sense left out, as the context within which the sentence is uttered makes plain the intended interpretation. These are radically incomplete strings, which have to be enriched in context by being assigned a structure in some sense provided by some antecedent structure, though, as we shall see it is not by any means a simple-minded copying of words that had previously been used. Take a scenario in which there's a smell coming from the kitchen, Alex rushes in to find Hugh with a smoking pan in response to which the following conversation takes place:

(28) Alex: Have you burned
Hugh: (interrupting): The buns? Very thoroughly. Myself, I think not.

The puzzle presented by exchanges such as (28) is that understanding Hugh's utterance, which is full of fragments – *the buns, very thoroughly*, and *myself, I think not* – involves syntactic and discourse properties simultaneously. Though the elliptical fragments are reconstructed as a process of interpretation, the output of such a building process displays structural properties exactly as do the sentences in which the structure of this interpretation might have been explicitly introduced. What Hugh has done in the exchange (28) is to take over and finish a partial structure which Alex has started out on. The proposition that, between them, they have constructed is that Hugh has burned the buns, even though Alex, by assumption, did not know what it was that was burning. It is that structure which, once established, Hugh goes on to extend, to convey the new information that he has burned the buns very thoroughly. What happens next is where there is use made of syntax mechanisms but not as defined over words. For, at this juncture, Hugh switches into using part of the structure he has just presented to create some new information. The expression with which he chooses to initiate the construction of this structure with is the word *myself*. To this, he adds the words *I think not*, thereby indicating that the information conveyed by the use of *myself* is going to have to be slotted into some structure to built up from what follows. This is exactly as in the topicalization structures of (27)–(30); except that the structure in question is never itself explicitly presented. Hugh presumes that the information is available in the context in some sense shared by the two of them, and indeed Alex has no difficulty in retrieving the information "burned," this time putting this together with the construal of Hugh's choice of word *myself* with the following *I think not*, to convey the information that Hugh thinks that he did not burn himself. So we have, in all, three assertions about some relation of burning, all in virtue of just one utterance of the word *burned*. Despite it being only Alex who has actually used the word *burned*, the construal of *myself* takes place exactly as though a syntactic operation had taken place, even though it can't have been the words used. We know this because if Hugh had explicitly used words in order to present the structure into which *myself* is to be slotted, he would have had to have switched away from Alex's use of the word *you*, saying instead: "Have I burned the buns? I have burned the buns very thoroughly. Myself, I think I have not burned." It is, of course, the nature of the words *I*

and *you* that they have to be switched in such an exchange, even though the content of what is being talked about, namely concern over whether Hugh is burned, remains constant.[22] But though humans interact in such exchanges with unerring fluency, it isn't trivial. What it shows is that in using elliptical constructions, people presume that what they say can be recovered from the content of what has already been said and that the phenomenon is not simply one of failing to repeat the precise words that were used. Thus, even without giving a formal step-by-step analysis of what is involved in this case, it is evident that the thoroughly context-dependent process of ellipsis construal has to be interwoven with supposedly syntactic processes such as reflexive-binding and long-distance dependency construal, so the construal of the reflexive in Hugh's reply in (28) can be seen as taking place in some sense after the process of ellipsis resolution.

Notice how, in this informal account, the speaker, just like the hearer, extends the representations under construction, in the speaker's case relative to further information they have in mind to convey, in the hearer's case only in their interpretation of what has been said to them. Moreover, this interaction is freely and seamlessly manipulated by interlocutors irrespective of which of the two roles they are playing. If, then, we take the syntactic mechanism of a speaker's grammar itself to be procedures for building up interpretation, this phenomenon is immediately predictable. Because the dynamics of interpretation growth is central to the concept of syntax defined in this framework, speakers must be seen as using this same mechanism. The only difference between building up structure for speaking and for hearing is that speakers must have some possibly marginally richer structure representing what they wish to convey which provides a check on what emergent structure can be constructed: The emergent structure must, for a speaker, be seen as compatible with this.

Strongly confirming this, such speaker/hearer shift of roles can take place across any supposedly grammar-internal dependency whatever:

(29) A: I need the dimmock.
B: The
A: dimmock, for breaking up the earth.

Furthermore, this phenomenon is controllable by children from a very young age, as witness the invariable pleasure that children very early on in the language-acquisition process display in engaging in joint performance of nursery rhymes such as (30):

(30) Carer (sings): Old McDonald had a farm ei-ei-o. And on that farm he had a
Child: cow.
Carer (sings): The cow goes
Child: Moo.

Notice in particular how the child has to produce a noun, when the adult stops with a determiner.

Neither of these phenomena should be dismissed as performance aberrations, or rhetorical tricks at the periphery of language. They display what is part of every fluent language user's capacity. It is notable that if, more conservatively, we persist in retaining a conception of grammar as inducing structure only as inhabited by complete strings, insisting, because the grammar is sentence-based, that dependencies are definable only over complete sentences, the emergent acquisition of parsing and production through interchangeability of language roles as speaker and as hearer from the very first stages is little more than a mystery. To the contrary, with a shift into seeing grammar as a system of procedures for interpretation growth, the phenomenon is expected, with interchangeability of roles anticipated from the outset of the acquisition process.

5 Summary

This chapter started by surveying the apparently acute problems facing any general account of communication. It set out the different types of separation that have been imposed on the distinction between linguistic knowledge and language use; and argued that, though there is a distinction between encoded and nonencoded information, nevertheless, there needs to be full integration of the two types of processing. It then sketched the two primary different sets of criteria for choosing the intended interpretation – the Gricean route through conformity to maxims governing conversational behavior, and the relevance theory route through a choice driven by balancing effort and effect. It then presented some of the evidence that words should be seen as procedures for interpretation, rather than having some fixed descriptive content. And, finally, it introduced one way in which the interpretation of topicalization structures and pronouns can both be explained in terms of the incremental building up of tree structure corresponding to interpretation. On the view that is emerging from this much more dynamic, use-oriented perspective on linguistic description, all words provide a set of procedures relative to which both hearers and speakers progressively construct a structure corresponding to the proposition expressed. The encoded specifications intrinsic to language are defined explicitly as the driving force in this incremental process of building up interpretations from a natural language sequence of words. Linguistic and nonlinguistic processes of interpretation, nevertheless, freely interact in determining what proposition or set of propositions is expressed by a given string of words. The only externally imposed restriction is that linguistic input, being an encoded set of instructions on structure-building, cannot be set aside. The linguistic knowledge that we have as users of the language is the encoded input which the individual language provides to enable the structural dynamics of the interpretation process to take place. Pragmatic principles are the general cognitive principles that enable us to enrich information by general reasoning strategies, and to make choices between alternative structures as the interpretation is progressively established. This strongly suggests that modeling the process of communication itself provides a basis for explaining what it means to know a language (= competence – see Chapter 13). This view is a departure from the view that linguistic ability should be a body of knowledge which is quite independent of whatever principles determine how language is used (= performance). On this view, natural language ability is, rather, a capacity for natural language processing. The current focus of debate in deciding on these views is barely begun, but at the center, as it always has been, is the status of our capacity for language. Is it, given the ability to set down a long-term store associated with words, a static store of knowledge relative to which pragmatic principles of use determine the apparently conflicting uses to which it can be put? Or is it the possession of a capacity for dynamically projecting structures which correspond to interpretation for a given piece of language uttered? At the present time, these remain questions over which researchers struggle to reach the most revealing answer, and much of what I have introduced in the last section of this chapter would be fiercely contested by many. One thing we can be sure of, however, is that the study of the way people use language is a central preoccupation of linguistic study.[23]

NOTES

1 This was once said to me as a fervent (and much appreciated) compliment!

2 What a user of a language knows about her language is called a *grammar*. In a grammar there are facts about the sounds used to build words (*phonology*), facts about how words are arranged to form sentences (*syntax*), and facts about those aspects of the meaning of words / sentences that are integral properties of those words / sentences (*semantics*) (see Chapter 10 on phonology; Chapter 13 on syntax; Chapter 14 on semantics).

3 A very useful collection which provides readings in all the major topics in pragmatics is Davis (1991).

4 Sentences in sequence are often used, contrarily, to imply events in sequence:
(i) I finished my PhD. I got it published. And I became pregnant.

5 It has long been recognized by both linguists and philosophers that pronouns such as *I* and *you* can only be interpreted relative to entities in the discourse context. In this connection, cf. the papers of Roman Jakobson writing in the 1950s, who referred to them as *shifters* (in Waugh and Halle 1984). In the philosophical literature, such pronouns are called *indexical pronouns*, cf. Perry (1979, 1993). However the recognition of the extent to which interpretation is dependent on context was not widely recognized until the early 1980s with the work of Kamp (1981), Barwise and Perry (1983), Sperber and Wilson (1995).

6 An important assumption here is that all cognitive activity involves the construction or manipulation of internal cognitive representations. We can, for example, only see that rose bush in the garden in virtue of setting up some internal representation of that rose bush (we don't have rose bushes in our heads – not even those of us who are obsessed with rose gardening). This view is known as representationalism and in its current form is largely due to Fodor, who has argued for a so-called language of thought (cf. Fodor 1975, 1981, 1983, 1998). What is controversial for linguistics is the relation of such representations to representations in linguistic description. There are substantial differences of opinion, for example, over whether any level of representation is required in modeling interpretation in language other than that articulated within syntax. This is a debate which has rumbled on in different forms for at least 30 years, cf. Katz (1972); Lewis (1972); Kamp (1981, 1996); Groenendijk and Stokhof (1991); Sperber and Wilson (1995); Kamp and Reyle (1993); Dekker (1996); Kempson, Meyer-Viol, and Gabbay (1997); Carston (1998).

7 A proposition needs to be complete only to the extent that it is one for which inference can be defined. So for example, absolute precision as to time is generally not required (I would bet that anyone reading this chapter has forgotten the exact date of Princess Diana's funeral). Given the nature of the conversation, suggesting a recent event, I have recorded the time variable as being "Saturday." The date was in fact Saturday September 6, 1997.

8 Contrast (1) with (i) which is inconsistent:
(i) I never cook anything but I make an omelette at the weekends.

9 The characterization of metaphor through the supposed violations of the maxim of quality is not entirely unproblematic. See Sperber and Wilson (1982) for a critique of Grice's theory of conversation and Wilson (1995) for a critique of the problems imposed by the maxim of quality characterization of metaphor. See also Section 4.1.

10 An anaphoric expression is one which can only be understood by reference to some other representation made available during the interpretation process. Pronouns are the central type of case, but words such as *the*, *this*, *that* are also anaphoric.

11 The qualification of "nontrivial" is important as from a purely formal perspective "P" implies "P and P," "P and P and P …"; P also implies "P or Q" for arbitrary Q, etc. A nontrivial inference is one which cannot be drawn from considerations of P alone.

12 The modification of maximal relevance to optimal relevance for the particular activity of utterance interpretation is because of the imposed task of recovering the speaker's intentions. Some utterance may be extremely relevant in terms of the hearer's own privately held assumptions but nonetheless not be the intended interpretation because the speaker couldn't possibly have had access to the assumptions that led to these implications. Paranoia is a good example of this, where one's private fears and anxieties are so easily retrievable that they constantly lead to inferential effects of one sort or another, but fortunately most of us, at least some of the time, realize that these private worries are not a reliable basis for recovering what speakers have intended to convey to us.

13 In logic, one is taught that from an inconsistency any proposition can be derived, so there is a trivial inference from "P and not-P" to Q for arbitrary Q. As with the earlier trivial inferences, these have to be explicitly debarred.

14 There is no fi xed choice of premises which must constitute the context, so any choice of premises will do as long as it licenses inferential effects associated with being at Princess Diana's funeral.

15 There are several alternative accounts of the way in which construal of language depends on the context in which it is uttered. In some of these, there is no commitment to any form of representation. Amongst these is Searle, who argues for a concept of *Background* relative to which language is interpreted, without any commitment to mental representations. Cf. Searle (1983, 1995).

16 The Speech Act theory of meaning, which was originally articulated as a theory of meaning for natural language under the slogan "Meaning

is Use," played an important part in the development of pragmatic theory. Language was explained in terms of different kinds of actions that can be carried out by the use of language. Primary distinctions were made between *locutionary acts* (what the speaker said to the hearer), *illocutionary acts* (what a speaker does by performing such locutionary acts – e.g., baptizing, promising, threatening), and *perlocutionary acts* (the effect a speaker has on the hearer – e.g., persuading, frightening).

17 This view has been consistently championed by Atlas in connection with negation (cf. Atlas 1977, 1989).

18 This is in contrast to Relevance Theory, where words in general are defined as providing concepts. Only a restricted subset of words provide what is called *procedural meaning*; words such as pronouns, and also connectives such as *but* (see Carston 2002).

19 The *lexicon* is the part of a grammar where all information about individual words is stored. Entries in the lexicon are generally referred to as *lexical items* but the simpler notion of *word* is sufficient for our present purposes. See Chapter 12.

20 Within this perspective, sequences such as relative clauses are analyzed as associated with a structure external to the matrix clause they modify, an analysis for which there is good reason within a perspective in which syntactic structure constitutes representations of content (rather than structure inhabited by words in sentences whose beginning and end are dictated by writing conventions). See for example the narrative flavor of (i) in which all clauses are joined together by a relative pronoun, the flavor of the narrative being much like the fully independent sentences of (ii):

(i) John ignored Mary, who complained to his mother, who duly reprimanded her son, who then apologized.

(ii) John ignored Mary. She complained to his mother, and she duly reprimanded her son. He then apologized.

21 The issue of ellipsis remains controversial. There are analyses of ellipsis which are purely syntactic (Fiengo and May 1994, and others), analyses which are purely semantic (Hardt 1993; Crouch 1995) or some mixture of the two (Kehler 1995; Lappin 1996). Cf. Kempson (1995); Kempson, Meyer-Viol, and Gabbay (1999) for an account of ellipsis in terms of how interpretation is inferentially built up as a propositional structure. For a representative selection of views, cf. Lappin and Benmamoun (1999).

22 This type of example is problematic for purely syntactic accounts of ellipsis, requiring a concept of "vehicle change" (cf. Fiengo and May 1994; Lappin 1996).

23 I have been helped in the redrafting of this chapter by detailed comments from Eleni Gregoromichelaki and the two editors Mark Aronoff and Janie Rees-Miller.

REFERENCES

Atlas, J. D. (1977). Negation, ambiguity, and presupposition. *Linguistics and Philosophy*, 1, 321–36.

Atlas, J. D. (1989). *Philosophy without Ambiguity: A Logico-Linguistic Essay*. Oxford: Clarendon Press.

Asher, N. and Lascarides, A. (2003) *Logics of Conversation*. Cambridge: Cambridge University Press.

Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Oxford University Press.

Bach, K. and Harnish, R. (1979). *Linguistic Communication and Speech Acts*. Cambridge, MA: MIT Press.

Barwise, J. and Perry, J. (1983). *Situations and Attitudes*. Cambridge, MA: MIT Press.

Blakemore, D. (1987). *Semantic Constraints on Relevance*. Oxford: Blackwell.

Blakemore, D. (1989). Denial and contrast: a relevance theoretic analysis of *but*. *Linguistics and Philosophy*, 12, 15–38.

Blakemore, D. (1992). *Understanding Utterances*. Oxford: Blackwell.

Brown, P. and Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Cann, R., Kempson, R., and Martin, L. (2005). *The Dynamics of Language*. Oxford: Elsevier.

Carston, R. (1988). Implicature, explicature and truth-theoretic semantics. In R. Kempson (ed.), *Mental Representation: The Interface between Language and Reality* (pp. 155–82). Cambridge: Cambridge University Press.

Carston, R. (1998). Pragmatics and the explicit implicit distinction. PhD dissertation, University College London.

Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Cohen, L. J. (1971). The logical particles of natural language. In Y. Bar-Hillel (ed.), *Pragmatics of Natural Language* (pp. 50–68). Dordrecht: Reidel.

Crouch, R. (1995). Ellipsis and quantification: a substitutional approach. *Proceedings of the Seventh European Chapter of the Association of Computational Linguistics* (pp. 223–26). San Francisco, CA: Morgan Kaufman.

Davis, S. (ed.) (1991). *Pragmatics: A Reader*. Oxford: Oxford University Press.

Dekker, P. (1996). Reference and representation. MS, University of Amsterdam.

Fiengo, R. and May, R. (1994). *Indices and Identity*. Cambridge, MA: MIT Press.

Fodor, J. A. (1975). *The Language of Thought*. New York: Crowell.

Fodor, J. A. (1981). *Representations*. Hassocks: Harvester Press.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon Press.

Gazdar, G. (1979). *Pragmatics*. New York: Academic Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (eds), *Syntax and Semantics*, Vol. 3. *Speech Acts* (pp. 41–58). New York: Academic Press.

Grice, H. P. (1989). *Studies in the Ways of Words*. Cambridge, MA: Harvard University Press.

Groenendijk, J. and Stokhof, M. (1991). Dynamic predicate logic. *Linguistics and Philosophy*, 14, 39–100.

Hardt, D. (1993). Verb phrase ellipsis: form, meaning and processing. PhD dissertation, University of Pennsylvania.

Heim, I. and Kratzer, A. (1998). *Semantics in Generative Grammar*. Oxford: Blackwell.

Kamp, H. (1978). Semantics Versus Pragmatics. In F. Guenthner and S. Schmidt (eds), *Formal Semantics and Pragmatics for Natural Languages* (pp. 255–87), Dordrecht: Kluwer.

Kamp, H. (1981). A theory of truth and semantic interpretation. In J. A. G. Groenendijk, T. M. V. Janssen, and M. J. B. Stokhof (eds), *Formal Methods in the Study of Language* (pp. 277–322). Amsterdam: Mathematisch Centrum.

Kamp, H. (1996). Discourse representation theory and dynamic semantics: representational and non-representational accounts of anaphora. MS, University of Stuttgart.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. Dordrecht: Kluwer.

Katz, J. (1972). *Semantic Theory*. New York: Harper and Row.

Kehler, A. (1995). Interpreting cohesive forms in the context of discourse inference. PhD dissertation, Harvard University.

Kempson, R. (1988a). The relation between language, mind and reality. In R. Kempson (ed.), *Mental Representations: The Interface between Language and Reality* (pp. 3–25). Cambridge: Cambridge University Press.

Kempson, R. (1988b). Logical form: the grammar–cognition interface. *Journal of Linguistics*, 24, 393–431.

Kempson, R. (1995). Ellipsis in a labelled deduction system. *Bulletin of Interest Group in Pure and Applied Logic*, 3, 489–526.

Kempson, R., Meyer-Viol, W., and Gabbay, D. (1997). On representationalism in semantics: a dynamic account of who. In P. Dekker, M. Stokhof and Y. Venema (eds), *The Proceedings of the 11th Amsterdam Colloquium* (pp. 193–9). Amsterdam: University of Amsterdam.

Kempson, R., Meyer-Viol, W., and Gabbay, D. (1999). VP ellipsis: towards a dynamic structural account. In S. Lappin and E. Benmamoun (eds), *Fragments: Studies in Ellipsis and Gapping* (pp. 227–90). Oxford: Oxford University Press.

Kempson, R., Meyer-Viol, W., and Gabbay, D. (2001). *Dynamic Syntax: The Flow of Language Understanding*. London: Blackwell.

Lappin, S. (1996). The interpretation of ellipsis. In S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory* (pp. 145–76). Oxford: Blackwell.

Lappin, S. and Benmamoun, E. (eds) (1999). *Fragments: Studies in Ellipsis and Gapping*. Oxford: Oxford University Press.

Lascarides, A. and Asher, N. (1993). Temporal interpretation, discourse relations, and commonsense entailment. *Linguistics and Philosophy*, 16, 437–93.

Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. (1987). Minimization and conversational inference. In J. Vershueren and M. Bertuccelli-Papi (eds), *The Pragmatic Perspective* (pp. 61–129). Amsterdam: John Benjamins.

Levinson, S. (1996). Three levels of meaning. In F. Palmer (ed.), *Grammar and Meaning* (pp. 90–115). Cambridge: Cambridge University Press.

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Lewis, D. (1972). General semantics. In G. Harman and D. Davidson (eds), *Semantics of Natural Language* (pp. 169–218). Dordrecht: Reidel.

Neale, S. (1996). Paul Grice and the philosophy of language. *Linguistics and Philosophy*, 19, 509–59.

Perry, J. (1979). The problem of the essential indexical. *Nous*, 13, 93–21.

Perry, J. (1993). *The Problem of the Essential Indexical and Other Essays*. Oxford: Oxford University Press.

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.

Searle, J. (1995). *The Construction of Social Reality*. London: Penguin.

Sperber, D. and Wilson, D. (1981). On Grice's theory of conversation. In P. Werth (ed.), *Conversation and Discourse* (pp. 155–78). London: Croom Helm.

Sperber, D. and Wilson, D. (1982). Mutual knowledge and reference in theories of comprehension. In N. Smith (ed.), *Mutual Knowledge* (pp. 61–121). Oxford: Blackwell.

Sperber, D. and Wilson, D. (1995). *Relevance: Communication and Cognition*, 2nd edn. Oxford: Blackwell.

Strawson, P. (1952). *Introduction to Logical Theory*. London: Methuen.

Walker, R. (1975). Conversational implicatures. In S. Blackburn (ed.), *Meaning, Reference and Necessity* (pp. 133–81). Cambridge: Cambridge University Press.

Waugh, L. and Halle, M. (eds) (1984). *Roman Jakobson 1896–1982: Russian and Slavic Grammar*. Berlin: Gruyter.

Wilson, D. (1995). Is there a maxim of truthfulness? *UCL Working Papers in Linguistics*, *7*, 197–214.

Wilson, D. and Carston, R. (2007). A unitary approach to lexical pragmatics: relevance, inference, and ad-hoc concepts. In N. Burton-Roberts (ed.), *Pragmatics* (pp. 230–259). London: Palgrave.

Wilson, D. and Sperber, D. (1988). Mood and the analysis of nondeclarative sentences. In J. Dancy, J. Moravcsik, and C. Taylor (eds), *Human Agency: Language, Duty and Value* (pp. 77–101). Stanford, CA: Stanford University Press.

21 Discourse Analysis*

AGNES WEIYUN HE

> *"We die. That may be the meaning of life. But we do language. That may be the measure of our lives."*
>
> —Toni Morrison (Nobel Lecture, 1993)

To imagine a world without discourse is to imagine a world without language and therefore to imagine the unimaginable. We get through our day exchanging various oral and written language (or, talk and text), delivered with specific pitches, volumes, facial expressions, gestures, emoticons. We live by *doing language*, or *languaging* or *discoursing*, not in discrete audio or visual units but in connected sound waves and orthographic forms to which we assign meaning on the basis of our past experience with them and on the basis of the situations in which these waves and forms are used. Discourse analysis is concerned with the contexts in and the processes through which we use oral and written language with specific audiences, for specific purposes, in specific settings. Many years ago, a well-known linguist, William Labov, observed that the term "sociolinguistics" is redundant – to him language is inherently social and so should be linguistics, the study of language. The same analogy may be drawn about discourse analysis: one cannot understand language fully without looking at language use.

1 What is Discourse? A Preliminary Characterization

The term "discourse" is used by many in very different senses, some having little to do with language. For instance, the term "globalization discourse" or "anti-terrorism discourse" as used in the media and some of the social sciences today often refers not to language use of those who practice globalization or anti-terrorism but to the ideologies and belief systems generated therein. In fact, studies of globalization or anti-terrorism discourse in political science or history, for example, may not pay any attention to language at all. Those studies are associated with the work of the French social theorist and philosopher Michel Foucault, who describes discourses as ways of framing the world. To differentiate discourse studies which do not focus on language and those which do, some researchers have made the distinction between the big D, which concerns general ways of viewing the world and general ways of behaving (including speaking), and

* The writing of this chapter was in part supported by a Guggenheim Fellowship. I am solely responsible for the statements made and the views expressed.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

the small d, which concerns actual, specific language use. Here, we are concerned with discourse as situated language use, or the small d, if you will.

What makes discourse analysis stand out as a discipline independent from, although intricately interwoven with, other domains of linguistic inquiry is that, more than any other domains of linguistics, discourse analysis emphasizes that language is not merely a self-contained system of symbols but more importantly a mode of doing, being, and becoming. Compared with other domains of linguistics which have a specific/definable scope of inquiry (e.g., phonology on sound systems, pragmatics on rules governing information processing in the mind), discourse analysis's concern may seem far broader and therefore more elusive: discourse analysis seeks to describe and explain linguistic phenomena in terms of the affective, cognitive, situational, and cultural contexts of their use and to identify linguistic resources through which we (re)construct our life (our identity, role, activity, community, emotion, stance, knowledge, belief, ideology, and so forth). Essentially, it asks why we use language the way we do and how we live our lives through the choices we make in language.

To this end, discourse analysis draws upon different disciplinary insights from linguistics, anthropology, psychology, sociology, communication studies, and education research. Discourse analysts insist on the use of naturally occurring, authentic language data (as opposed to invented or introspective data). However, in spite of the shared global aim, different discourse analysts may focus on vastly different aspects of communication, draw upon divergent analytical traditions, and resort to sometimes incommensurable theories and methodologies. The main purpose of this chapter is not to systematically trace the intellectual genesis or diversity of discourse analysis or to review all approaches and methods used in discourse analysis. Readers interested in these issues are referred to Schiffrin (1993), Ochs, Schegloff, and Thompson (1996), Duranti (1997), Wodak and Meyer (2009), Schiffrin, Tannen, and Hamilton (2003), Blommaert (2005), Widdowson (2007), Johnstone (2008), and Strauss and Feiz (2013). My objective here is to address in simple terms the following question: What is discourse analysis and what do discourse analysts do?

As mentioned before, discourse analysis focuses on language as not merely systems of acoustic and orthographic symbols and rules for sequencing words or inferring meaning, but rather on language use motivated by real communicative needs and language as a means through which we accomplish various actions and interactions. Broadly speaking, discourse research can be divided into two major types of inquiry: (1) why some but not other linguistic forms are used on given occasions; and (2) what are the linguistic resources for accomplishing various social, affective, and cognitive actions and interactions. Below we will consider both emphases.

1.1 Communicative motivations for the selection of linguistic forms

One set of questions that discourse analysts are concerned with has to do with the communicative motivations for the selection of linguistic forms (e.g., Fox 2007). Unlike formal linguists (or autonomist linguists) who believe that language is a self-contained system, discourse linguists maintain that language is inseparable from other aspects of our life and that the selection of linguistic forms should be explained in terms of authentic human communicative needs (i.e., social, interactional, cognitive, affective). This position is compatible with and is inspired by insights from a number of different disciplines including anthropology, cognitive science, functional linguistics, psycholinguistics, philosophy, and sociology.

The following is a transcript of the beginning of an actual face-to-face meeting between a university student Susan (S) and her academic advisor Neil (N). The meeting was audio-recorded by me as part of a project on institutional discourse (He 1998). S intends to go to medical school and wishes to find out which major will give her the best chance of being accepted by medical school. The "Helen" mentioned in their talk is another academic advisor. I will use this segment

and segments related to it to illustrate some of the main issues which concern discourse analysts and some of the important analytical traditions which have influenced the practice of discourse analysis.

(1)
":" = prolonged syllable; "=" = latching, i.e., when the current speaker barely finishes, the next speaker begins; "CAPS" = stressed syllable; "(.)" = untimed micropause; "." = falling intonation; "," = rising intonation; "(0.0)" = silences roughly in seconds and tenths of seconds; "( )" = undecipherable hearing; "[" = onset of overlap.

```
001   N:   So:.
002   S:   All right um so,=
003   N:   =RIGHT now you are a math major.
004   S:   I AM a math I mean I TRANSferred as a math major.
005   N:   Ok.
006        (.)
007   N:   Oh (.) Probably PRE-math.
008   S:   Premath (.) that's right=
009   N:   =Ok,=
010   S:   =Lemme=
011   N:   =(  )look at your file.
012   S:   Trans- yeah I HAVE that.
013        (.8)
014   S:   Ok I'll FIND it in a second,
015   N:   All right=
016   S:   =But um (.2) see (.) um I: would like to go to (.)
017        med school,
018   N:   Uhuh,
019   S:   Ok, (.8) and uh when I (.2) when I was in the
020        orientation, (.) Helen told me that (.2) it's a
021        LOT better if I am a MATH major, (.) 'cause uh
022        medical schools they prefer math major people.
023        (.4) And I am not sure how that I mean I I
024        believed her THEN but NOW I've been talking to
025        [people
026   N:   [And NOW you DON'T believe her.
027   S:   Yeah I am NOT sure if that is the (.2) the RIGHT
028        thing or no:t.
029   N:   I would say um (.) I'm not as much of an expert (.)
030        about what happens to math majors (.) as Helen is.
031        She's (.2) doing research with what WHAT (.) has
032        happened to CU math majors and where they GO.
033        (.3)
034   N:   Uh but I'd say that certainly (.) medical school
035        doesn't CARE what your major IS.
036        (.8)
037   N:   Y=
038   S:   =Yeah that's what I heard.
039   N:   What they do care is (.2) uh did you take the
040        appropriate classes,  Do you have the: (.) the
```

```
041       grades for appropriate classes,  Do you have the
042       overall GPA do you have letters of recommendation
043       and so on so on so on.
044       (.5)
```

Looking at this transcript, discourse analysts will consider a number of things, centered around the question of why the speakers select the linguistic forms that they do in this particular moment. Below we discuss some of the concerns, not in any particular order.

1.1.1 Context One of the first questions discourse analysts ask is what is happening in this stretch of talk, who the participants are, where they are and why they are there. In other words, discourse analysts are quintessentially concerned with the *context* of language use. For discourse linguists believe that linguistic choices are not made arbitrarily but rather are systematically motivated by contextual factors. One influential model to describe context comes from a branch of linguistic anthropology known as ethnography of speaking or ethnolinguistics or ethnopragmatics (Duranti and Goodwin 1992; Duranti 1997, 2009). In this model, the range of parameters which need to be taken into consideration when characterizing context is summarized by the *SPEAKING* acronym (Hymes 1974), which stands for Situation (setting, scene), Participants (addressor and addressee), Ends (goals and outcomes), Act sequence (message form and message content), Key (tone, manner), Instrumentalities (channel, forms of speech), Norms (norms of interaction and norms of interpretation), and Genres (text types). Context as described in this way helps us understand what type of *speech events* or *speech activity* – a set of interpersonal relationships and a set of interpretative frames in relation to some communicative goal – we are dealing with. Given that we have in front of us an academic counseling encounter and not a job interview, for example, we will have certain expectations about topical progression, turn-taking rules, and outcome of the interaction as well as constraints on context. In this case, the Situation is an academic advisory meeting taking place in a university. Participants include the student and her advisor. The Ends for the advisor are to help the student choose the optimal path in her academic career while observing the university's rules and policies. The Act sequence entails problem presentation and problem solution. The Key is formal but friendly. The Instrumentality is spoken language. The Norms include the advisor withholding personal opinions and leaving decision making to the student. And the Genre is a face-to-face interview. Along the same line of ethnography of speaking but assuming a poststructuralist orientation, a more recent effort called "linguistic ethnography" (Creese 2008; Rampton 2006) aims to further fuse linguistic analysis with ethnographic analysis by scrutinizing the detailed interplay between language and culture, the patterned and dynamic nature of this interplay, and the emergent nature of meaning-creation in the making of context.

Another important tradition which has impacted the approach to context in discourse analysis is a school of functional linguistics called Systemic Functional Grammar (Eggins 2005; Halliday and Mathiessen 2014). In this theory, the notion of context owes its origin to the notion of *context of situation* developed by the British anthropologist Bronislaw Malinowski. In Malinowski's sense, the context of situation is indispensable for the understanding of utterances; in other words, a given utterance has no meaning unless it is accompanied by ethnographic analysis of situations within which interaction takes place. This notion was later elaborated and modified by J. R. Firth, a British linguist, who points out that a context of situation for linguistic work entails the following dimensions: (1) the relevant features of participants: persons, personalities, both verbal and nonverbal actions of the participants; (2) the relevant objects; and (3) the effect of the verbal action. When the idea is further developed by Halliday (1985), the leading figure in Systemic Functional Grammar, and his colleagues, context is represented as a complex of three dimensions: First, the *field* is the field of social action in which the discourse is embedded. Second, the *tenor* is the set of role relations among the relevant participants, including their

statuses and roles and their permanent and temporary relationships. And third, the *mode* is the role of language in the interaction. In this view, language is a system of choices, choices made on the basis of a contextual configuration which accounts for field, tenor, and mode.

In our example, what is taking place (Field) is seeking and giving advice concerning academic matters in an American university setting. It is a routine activity for many undergraduate students in US universities and colleges and their on-campus academic advisors. The relationship between the student and the advisor (Tenor) is a complex one: on the one hand, the advisor represents the university institution and implements the university's various rules and policies; on the other hand, the advisor is also an advocate of the student's interests. Language in this instance constitutes the advising activity itself (Mode) – it is through talk that problems are identified and solutions formulated.

1.1.2 Rhetorical goal On a more local level, some discourse analysts are concerned with the *rhetorical goal* of utterances. Consider lines 19–22 for instance:

```
019  S:  Ok, (.8) and uh when I (.2) when I was in the
020      orientation, (.) Helen told me that (.2) it's a
021      LOT better if I am a MATH major, (.) 'cause uh
022      medical schools they prefer math major people.
```

Is Susan's utterance here designed as a simple narration of a past event (that Helen told her something at some point) or is it a precursor to her counseling request (i.e., please tell me whether Helen is correct so that I can make up my mind about my major)? Does it function as an elaboration of what has been said before in line 16 "But I'd like to go to medical school" in which "but" already forecasts trouble? Or does it serve as a contrast to what has been said? These concerns fall within the domain of "rhetorical management" (Chafe 1980) or "rhetorical structures" (Mann, Mathiessen, and Thompson 1992; Taboada and Mann 2006) of information.

1.1.3 Speech act Discourse analysts following a philosophical tradition called Speech Act Theory will be asking what kind of *speech act* is Susan's utterance in these same lines and whether this act is accomplished through direct or indirect means. Speech Act Theory (Austin 1962; Searle 1969, 1979) says that language is used not only to describe things but to do things as well. For example, we use language to make requests (belonging to the class of directives), promises (commissives), or apologies (expressives). Utterances in instances like these (e.g., "But I would like to go to medical school") cannot be assessed in relation to truth and falsity; instead, they can only be assessed in relation to the necessary and sufficient conditions (known as "felicity conditions") for their accomplishment. Further, utterances act on three different levels: the literal level (locutionary act), the implied level (illocutionary act), and the consequence of the implied act (perlocutionary act). The locutionary meaning of what Susan says in lines 19–22 is thus what Helen told her. The illocutionary force is however subject to further determination: is it an indirect request for a second opinion? Or does it imply an invitation for Neil to ask further questions? Whichever the case, the perlocutionary upshot can be that Susan is taking advantage of the advisor-student relationship on this occasion to get Neil to do things without explicit requests. This line of inquiry leads to the analysis of language use by combining the analysis of the propositional content of utterances with their illocutionary force, which in turn permits us to draw inferences about the speaker's intentions and his/her inner world of beliefs, assumptions, desires, attitudes, stances, and so forth.

1.1.4 Scripts/plans Psychologically oriented discourse analysts will also be interested in these lines, but more from the viewpoint of how this utterance will be processed in the head of the listener. What sorts of *scripts*, *plans*, and *macrostructure* (van Dijk and Kintsch 1983; Schank and

Abelson 1977) are relevant and necessary for Neil to understand Susan's utterance as it is intended? For instance, Neil would need to have a knowledge of what often takes place in an advice-seeking and advice-giving encounter so that any narration of past events may be assessed in terms of their problem-ridden nature. He would also need to properly position Susan's utterance in the context of her immediate and overall goals as well as in the context of the overall structure of meetings of this kind (which often begin with problem presentation, followed by problem solution, etc.). Script is a term used to describe the knowledge that we have of the structure of stereotypical event sequences such as seeking and giving academic advice in a university institutional setting. If such knowledge can be described in a formal way with explicit rules and conditions, then we may have a theory of how humans process natural language; furthermore, we may be able to program and test that knowledge on a computer.

1.1.5 Referentiality Motivated by an interest in the organization of information and a concern with the relationship between language and body, space, and ritualized practice, discourse analysts have paid special attention to how participants refer to themselves, to each other, and to their world (Fox 1987; Hanks 1990). Let us take a look at how Helen is referred to in the data segment. Below is the subsection containing reference to Helen.

```
019  S:  Ok, (.8) and uh when I (.2) when I was in the
020      orientation, (.) Helen told me that (.2) it's a
021      LOT better if I am a MATH major, (.) 'cause uh
022      medical schools they prefer math major people.
023      (.4) And I am not sure how that I mean I I
024      believed her THEN but NOW I've been talking to
025      [people
026  N:  [and NOW you DON'T believe her.
027  S:  Yeah I am NOT sure if that is the (.2) the RIGHT
028      thing or no:t.
029  N:  I would say um (.) I'm not as much of an expert (.)
030      about what happens to math majors (.) as Helen is.
031  N:  She's (.2) doing research with what WHAT (.) has
```

We see that when Helen is first mentioned in this spate of talk, she is referred to by her name (line 20). When she is mentioned again in the same speaking turn by Susan, she is referred to as "her" (line 24). She is referred to in the same way by Neil in line 26. The choice between a proper noun "Helen" and a pronoun "her" (and other possibilities such as "that lady," "this other advisor") partly has to do with whether and how Helen has been mentioned before in prior talk and whether the entity "Helen" is recoverable or accessible from previous discourse or from the listener's existing knowledge. In other words, it has to do with the *information status* of "Helen" – whether "Helen" is *given* information or *new* information (also known as old vs. new, or shared vs. new, or known vs. unknown) (Prince 1981; Givon 1983). In line 20, since Helen has not been mentioned before, the choice of "she/her" is out of the question. And presumably since academic advisors know each other and often address each other on first name basis, it is appropriate and efficient for Susan to choose "Helen" among all possibilities. In line 24, since Helen has been introduced into the talk and since no other female person has been mentioned, "her" is the most efficient reference and cannot be mistaken for anyone else. In line 26, Neil shows that he really follows what Susan is saying by predicting what she will say next; "her" in this case, then, is intended and can only be interpreted to be coreferential with "her" in line 24. Some discourse analysts, sometimes known as text linguists (Halliday and Hasan 1976; de Beaugrande and Dressler 1981), are particularly interested in how referential forms help make a stretch of discourse *cohesive* in form and *coherent* in meaning.

We may then ask why in line 30 Neil refers to Helen by her name again, since he has already previously referred to her with a pronoun "her." Well, besides old vs. new, other factors come into play. Note that between Neil's two speaking turns which respectively begin at line 26 and line 29, there is an intervening turn by Susan (lines 27–28). And that turn is about "I" (Susan) and says nothing about Helen. When Neil speaks again in line 29, he speaks of himself in comparison with Helen. In other words, lines 19–28 constitute a problem-presentation phase; beginning from line 29, Neil starts giving advice. These two sections are two different *episodes* (Fox 1987; van Dijk and Kintsch 1983). Discourse studies have shown that at the beginning of a new episode, references are likely to be made in ways similar to how they are made for the first time.

1.1.6 Topicality and thematicity Another set of issues which concern how information is presented – what is an utterance about, what is the starting point of a message, what is the focus of a message – includes *topic* and *theme*. These are some of the most controversial concepts in discourse studies; to this day there remain substantial disagreements among discourse linguists as to whether these two notions are the same and what they entail.

Topic has been defined in terms of aboutness, i.e., the part of the utterance about which something is said. It has also been considered in terms of its grammatical and discourse functions. For example, in some languages such as Chinese, the element central to discourse is often placed initially preceding the utterance and assumes a certain grammatical role. (E.g. "小妹, 我早就认识她了. *Xiaomei, wo zao jiu renshi ta le*" (Xiaomei, I have known her for a long time) where "Xiaomei" specifies the topic.) These languages are called topic-prominent languages (Li and Thompson 1976).

The notion of Theme was originally developed by a European linguistic tradition known as the Prague School (Danes 1974; Mathesius 1975). Working with Slavic languages whose word order is more flexible than in languages such as English and which depends crucially on degrees of knownness/givenness of information, Prague School linguists developed what is called the *functional sentence perspective*. This states that word order has to do with how informative each element in the utterance is – *communicative dynamism*. A sentence begins with elements with the lowest dynamism and ends with those with the highest dynamism. Theme, in this framework, is the part of the utterance with the lowest degree of communicative dynamism. Influenced by the Prague School approach, linguists working within Systemic Functional Grammar take Thematic information to be information from which the speaker proceeds and Thematic organization to be a method of development of the text (Fries 1981, 2009; Hasan and Fries 1995).

Back to the interaction between Susan and Neil. We have seen that at the beginning of their meeting, Susan reported to Neil what Helen had told her about choices of major for medical school. At a later point, Neil comments as follows (presented clause-by-clause; indentation shows subclauses; initial elements up until the grammatical subject of the clause are in bold face):

(2)
Neil2:121-137
N: **y' see** there're there're two ways (.2)
you can read what she said.
One way is (.)
medical schools look at transcripts
and look for major
and they see math major
and they circle with a red pen
and they they add ten points to your your score or something
and they let you in more often.

```
(.3)
N:          The other thing is
            they look at your transcript
                  and look at your scores (.2)
            and (they look at) your MCAT
                  and look at your letters of recommendation(.2)
            they admit people
            and when you look at the people they admit,
                  there are more math majors than you expect.
      (1.0)
```

Here Neil outlines two ways of interpreting what Helen said with "One way" and "The other thing" at the beginning of his two main clauses. In each interpretation, he begins with "medical schools" or "they." A topical analysis says that, in this stretch of talk, "medical schools" is the entity about which something is said. A thematic analysis says that "medical schools" specifies the frame within which something is said. Hence Topic and Theme may (but do not necessarily) reside in the same elements, but they represent different meanings.

1.1.7 Sequential organization The selection and interpretation of linguistic forms in a given utterance have a lot to do with the *sequential context* of the utterance (Couper-Kuhlen and Selting 1996; Ford 1993; Ford, Fox, and Thompson 2002; Ochs et al. 1996; Schegloff 1979, 1996). For instance, where is the utterance positioned in interaction? Is this utterance opening a dialog? Or is it a response to some prior talk? Or is it a repair of some prior talk? Or is it part of a narrative episode? Discourse analysis in recent years has been profoundly influenced by a distinct approach to human interaction known as Conversation Analysis (Sacks 1992; Atkinson and Heritage 1984; Hutchby and Wooffitt 2008; Schegloff 2007; ten Have 1999). This analytical tradition has provided the most systematic and rigorous account of how we manage talk in interaction. Central to Conversation Analysis is the concept of *turn taking*, which can be described by a set of rules with ordered options that operate on a turn-by-turn basis as a locally, sequentially managed system. This system explains how speakers 'earn' their right to speak, how speaking rights are negotiated and interactionally managed, how the next speaker is selected, how overlaps occur and how they are resolved, and how speakers fix problems in comprehension and miscommunication. A turn is constructed with turn-constructional-units, which refer to sentential, clausal, phrasal, lexical, intonational, or pragmatic units (Sacks, Schegloff, and Jefferson 1974; Ford and Thompson 1996; Ford et al. 2002). The rules of turn-taking apply at the end of each turn-constructional-unit, which is called a transition-relevant-place.

I will introduce two more concepts from Conversation Analysis to prepare for the discussion of the following sample data segment. One of them is *adjacency pair* (Schegloff and Sacks 1973) – a sequence of two utterances adjacent to each other, produced by different speakers, ordered as a first part and second part, and typed, so that a first part requires a particular second part or range of second parts. A good example of an adjacency pair is a question-answer pair. The other concept is called *repair organization* (Schegloff, Jefferson, and Sacks 1977). When trouble in conversation occurs, it is noticed and then corrected, either by the party whose turn contains the source of trouble or by some other party. This sequence of trouble + initiation-of-correction + correction is known as a *repair trajectory*. Repair occurs when one party corrects his or her own talk or that of another party and can be accomplished in a number of ways.

Now let's consider the following segments (3) and (4) from the same meeting between Susan and Neil. Here they are discussing specific courses Susan has taken and can take the next semester. I wish to show through these two segments how Neil and Susan define and modify the meanings of the modal verb "can" through interaction.

(3)
"*italics*" = modal verbs
230 S: No. Uh I'm going to (.2) well I have a Fortran,
231 N: Uhuh,
232 S: That Helen said we *can* just=
233 N: =*should should should* accept that.
234 S: Right.

(4)
283 N: Ok. [you *can* swi- you *can* switch.
284 S: [and I figured it's a lot harder.
285 S: I *can*?=
286 N: =Y'know you *can* take 8A 6B 6C if you want.
287 That's fine. [No problem.
288 S: [Ok. Yeah.
289 (.)

In theory, "can" can mean a number of different things: it can index the speaker's knowledge or belief, it may imply necessity or logical possibility or probability, it may indicate interpersonal obligation, and so forth. In (3), Neil completes the turn that is initiated by Susan (lines 232–233). In the course of doing so, he replaces S's reported "can" (line 232) with "should" (233) and finishes S's turn, thereby interpreting S's report of what Helen said. S agrees with such an interpretation ("Right," line 234). Thus through Neil's completion and repair of Susan's turn and Susan's subsequent ratification, Neil and Susan jointly interpret "can" as used by Helen to mean "should."

(4) is a case involving S initiating repair (line 285) of N's turn (line 283), which results in N's repair in line 286. The interpretation of "can" in line 285 and line 286 should not be made arbitrarily; it should be made in relation to "can" in the original turn of line 283, as line 285 is a repair initiation of 283 and 286 is a repair (a specification) of line 283.

The above is not an exhaustive list of features to which discourse analysts attend. Neither is it an inventory list for anyone who does discourse analysis to follow. It only gives us an idea of some of the features which often draw the attention of discourse analysts. Depending upon their intellectual persuasion, different discourse analysts may focus on one or several of the above features. There is, however, no single study which considers all of them at the same time.

We have so far considered one set of questions which occupy discourse analysts: on given occasions, why do we select the linguistic form that we do? We have seen that the answer to this question is complex and multifaceted. Discourse analysts have sought to explain linguistic choices in terms of ethnographic contexts, knowledge structure, rhetorical organization, communicative intentions, textual organization, information management, and conversational sequential organization, among others. Next we turn to the other set of questions: How does language use contribute to and construct our social, cultural, intellectual, and emotional life?

2 Linguistic Resources for Doing and Being

As discussed previously, discourse linguists take a dialectical view of the relationship between language and other aspects of human life. On the one hand, the selection of linguistic forms is shaped by various contextual factors; on the other hand, the way in which we use language contributes to/constitutes other aspects of life. Thus our second set of questions is a corollary of the first set. In this section, I first sketch some of the topical areas of work along these lines. I then illustrate how analysis may be performed by returning to the interaction between Susan, the university student, and Neil, her academic advisor.

Some of the earlier work in discourse studies concerned how ***ordinary, everyday activities*** are accomplished linguistically. The tradition of Conversation Analysis, for instance, was developed in the late 1960s and 1970s and was motivated by a concern with the linguistic and interactional resources that we deploy in making sense of everyday life. Conversation, or talk-in-interaction, is considered the primordial means through which we conduct our social life and construct various social relationships (Schegloff 2007). Discourse work from early sociolinguistics looked at, for example, how syntactic structures reconstructed life experience in narratives (Labov 1972).

Since the late 1980s, there has been a surge of research interest in how ***institutional, professional activities*** are carried out linguistically (e.g., Antaki, Barnes, and Leudar 2005; Arminen, Auvinen, and Palukka 2010; Barnes 2007; Cazden 1988; Drew and Heritage 1992; Gunnarsson, Linell, and Nordberg 1997; Peräkylä et al. 2008; Seedhouse 2004; Young 2009; Young and He 1998). How do doctors talk to their patients, therapists to their clients, advisors to their advisees, business people to their counterparts, teachers to their students, prosecutors to their witnesses, interviewers to their interviewees, etc., etc.? What kinds of speech exchange system (e.g., a particular turn-taking pattern; storytelling) are characteristic of the particular speech event? What are some of the salient lexico-grammatical forms (e.g., modal verbs such as "can" and "must"; generic uses of personal pronouns such as "you" or "they") used in these events? How does the way in which the participants use language reenact, maintain, or alter their institutional roles and identities? How are institutional activities accomplished through verbal and nonverbal interaction?

Furthermore, discourse analysts have undertaken to examine how ***identities*** are reconstructed linguistically. Instead of treating the language user's identity as a collection of static attributes or as some mental construct existing prior to and independent of human actions, discourse analysts approach identity as something dynamic which is continually emerging and which identifies what a person becomes and achieves through ongoing interactions with other persons and objects (Block 2007; Bucholtz and Hall 2005; He 2011; Ochs 1993). Along this line, gender identities (M. Goodwin 2006; Tannen, Kendall, and Gordon 2007; Speer and Stokoe 2011), ethnic/cultural identities (Rampton 1995; Scollon and Scollon 1981; Wierzbicka 1992; Mendoza-Denton 2007) have been examined as they evolve together with language use.

In a similar vein, discourse analysts have pursued the role of language in the (re)construction of ***ideology*** (Schaeffner and Holmes 1996; Pütz, Neff-van Aertselaer, and van Dijk 2004; Schieffelin 2007), ***epistemology*** (Whalen and Zimmerman 1990; Heritage and Raymond 2005; Kidwell 2009; Koole 2010), ***emotion*** (Capps and Ochs 1995; Wilce 2009), ***cognition*** (C. Goodwin 1994, 2003; Heritage 2005), and, among other facets of life, ***cultural ecology*** (Barton 2007; van Lier 2004; Kramsch 2002; Kramsch and Whiteside 2008).

Last but not least, research from a branch of linguistic anthropology known as Language Socialization has been dedicated to the scrutiny of the impact of language use on the ***socialization*** of values, norms, and other sociocultural knowledge. The focus is on the organization of communicative practices through which novices (e.g., children) learn to become competent members of their communities (e.g., Heath 1983; Ochs 1988; Schieffelin 1990; Duranti, Ochs, and Schieffelin 2011). This body of research examines audio-/video-recorded, carefully transcribed, recurrent socialization activities involving experts (e.g., caregivers) and novices (e.g., children) and explores the impact of the verbal and nonverbal details of interaction on the construction of social and cultural ideologies that define a community.

Space does not permit me to include more topical areas of discourse work or to cite more studies for illustrative purposes. We can safely say that with our life and our world becoming increasingly rich and diverse, the possibility and potential for discourse analysis is unlimited. Let me next return to the data segment about Susan and her academic advisor Neil to illustrate how analysis of language data vis-a-vis some of the concerns sketched above may be carried out.

2.1 Roles and identities

The first order of business when the participants sit down in the counseling office is to establish who and what they are to each other with regard to the counseling encounter. This may sound superfluous as clearly both participants already have pre-existing respective identities as "counselor" and "student" defined by the university. But these pre-existing titles do not specify what kind of counselor or student they are, nor do they provide concrete guidance as to how, for example, one student's problem might be treated differently from another's or how one counselor's advice might be taken differently from another's. Despite their generic roles and identities, the participants still need to establish their attributes specific to their meeting – the student in terms of his/her academic conditions which occasioned his/her visit and the counselor in terms of his/her authority and expertise with respect to the student's conditions.

In the data segment in Section 1.1, how to categorize the student in terms of an academic major is the focus of talk in the initial minute. Is she a math major or a pre-math major? Is she a committed math major or a math major by virtue of the fact that she transferred to CU as a math major? How do the participants arrive at a shared understanding? Why do the participants choose academic major as an attribute relevant and important to the beginning of their encounter? How is this categorization to affect subsequent talk and subsequent counseling activities? Is the counselor a knowledgeable and competent one?

2.2 Activities and tasks

In addition to establishing occasion-specific identities, the participants also need to determine and agree upon the task and purpose of their meeting. What is the counseling problem? How is the problem presented, identified, and formulated? What role does each participant play in the identification and formulation of the counseling problem? When the student makes an appointment with the academic counseling office, he/she is asked what his/her needs are. The university representative who schedules the appointment summarizes the student's response on the basis of the former's understanding of what the problem is or could be. With that summary he/she (the representative) then fills in the slot "purpose of visit" on the appointment slip. Hence what is stated on the appointment slip as "purpose of visit" reflects how the person who schedules the appointment characterizes the problem, a characterization often inevitably too distilled to be informational and sometimes unfortunately too speculative to be accurate.

In the case of Susan, "choosing a major" was put down as purpose of visit on the appointment slip. This could mean a number of different things. It could be that the student has not declared a major and is in need of help with choosing one. It could be that the student has not officially declared a major but has already made up her mind about which major to choose and is therefore in need of help with technical procedures to get into the major. It could be that the student already has a major but for one reason or another is contemplating changing to a different major. It could also be that the student has a number of concerns, the top one on the list being matters related to choosing a major. Each of the above scenarios presents a different set of tasks and activities – different forms to fill, different kinds of information to discuss, and different ways of talking and interacting. Although the student's record file can supply certain information (e.g., whether or not the student has officially declared a major), which scenario each specific case is can only be determined through the actual encounter between the student and the counselor.

It turns out that Susan already has a major. How do the counselor and student unpack the stated purpose of visit "choosing a major"? Note that nowhere in the above segment can we find utterances such as "I am here because I would like to…" or "Tell me specifically what I can help you with." How then do the participants come to a shared orientation to reconsidering majoring in mathematics as their task?

For ease of reference, let me reproduce the first 8 lines below:

```
001    N:    So:.
002    S:    All right um so,=
003    N:    =RIGHT now you are a math major.
004    S:    I AM a math I mean I TRANSferred as a math
             major.
005    N:    Ok.
006          (.)
007    N:    Oh (.) Probably PRE-math.
008    S:    Premath (.) that's right=
```

We see that in the very beginning Neil identifies Susan in terms of her major (line 3, "you are a math major") and displays his orientation to the service request; his "RIGHT now" (line 3) casts being a math major as S's temporary identity and forecasts a change of major, which will be the focus of discussion. Subsequently, Susan corrects Neil's categorizing her as a math major through a self-repair of her own utterance (line 4, "I AM a math I mean I TRANSferred as a math major"). By changing from "I AM," indicating a present state of being, to "I TRANSferred," indicating a past action, Susan focuses on the process of how she became a math major. She also displays a lack of commitment to being a math major, and thereby converges with Neil's understanding (line 3) that their encounter is to be about a change of major. In what follows, Neil revises "math major" into "PRE-math" (line 7), an official university category which on this occasion also echoes and reinforces Susan's tentativeness displayed in her previous turn. Susan then acknowledges that Neil's categorization of her is accurate (line 8). Hence, before the first seven utterances are completed, a shared orientation to the student's identity and the institutional task at hand has been constituted.

Susan's identity, in this instance her academic major, is co-constructed in this case to scaffold the counseling problem (i.e., choosing a major); the selection of "math" or "pre-math" as an identificatory category is bound to what the participants collectively take this particular counseling encounter to be about. It can also be seen that the establishing of institutional identity is a reciprocal process; the establishment of the student's identity also establishes that of the counselor. Recall that Susan produces her initial correction of Neil as a self-repair of her own utterance ("I AM a math I mean I TRANSferred as a math major," line 4). With this self-correction, it appears that she is correcting herself and not the advisor, a strategy by which she collaborates in constituting the advisor's expert role. Though Neil mitigates his subsequent correction by using an adverb of uncertainty ("probably") in "probably PRE-math," this choice of modifier in fact enhances his expertise by invoking a large body of experience from which he can generalize. The advisor thus demonstrates his ability to make judgments on the probability of facts which are not explicit. By situating the identity of Susan's major within the specific counseling task at hand and in relation to the advisor's expert role, we are now able to view the identity of the student not only as embedded in the academic counseling context, but also actively contributing to the context.

2.3 Knowledge and stances

Equally important are the institutional knowledge the participants exhibit and the affective stances they project. How does Susan present her understanding that being a math major may pose a problem for her to later attend medical school? Does she appear to be certain or tentative? In either case, what does she do linguistically to help her construct that appearance? What is the function of quoting Helen, another advisor? What role does Neil

play while the student is fleshing out the purpose of her visit? How does Neil address her concern? Does he provide a straightforward answer? Is he assertive, direct, cautious, and/or empathetic?

In more general terms, how do the participants come to know what they know? How do they assign responsibility of what they say and what they experience? How do they position themselves in relation to each other and to the university institution? How does the counselor manage the dual role of being an institutional gatekeeper and an advocate for the student's interests?

Again, for ease of reference, here are lines 16–33:

```
016  S:  =But um (.2) see (.) um I: would like to go to (.)
017      med school,
018  N:  Uhuh,
019  S:  Ok, (.8) and uh when I (.2) when I was in the
020      orientation, (.) Helen told me that (.2) it's a
021      LOT better if I am a MATH major, (.) 'cause uh
022      medical schools they prefer math major people.
023      (.4)  And I am not sure how that I mean I I
024      believed her THEN but NOW I've been talking to
025      [people
026  N:  [NOW you DON'T believe her.
027  S:  Yeah I am NOT sure if that is the (.2) the RIGHT
028      thing or no:t.
029  N:  I would say um (.) I'm not as much of an expert (.)
030      about what happens to math majors (.) as Helen is.
031      She's (.2) doing research with what WHAT (.) has
032      happened to CU math majors and where they GO.
033      (.3)
```

Here Susan reports that Helen has stated that medical schools accept a higher percentage of math majors. In the report, Susan carefully retains her own attitude separate from Helen's through her distinct temporal and modal choices. What Helen reportedly said is encoded with straightforward tense markers only (lines 19–22); whereas the student's own commentary is encoded with hedges ("not sure" in lines 23 and 27) which help construct a sense of doubt and uncertainty in contrast to the certainty and truthfulness of Helen's speech. Thus Susan portrays the account of the problem as certain and truthful and her own attitude as doubtful and uncertain, which in turn warrants Neil's attention, alignment, and subsequent advice. And Neil is not merely a passive recipient of Susan's report. Rather, he actively anticipates her account (line 18), sympathetically collaborates with her in her account (line 26), and cautiously provides his assessment of the situation without discrediting Helen, his colleague (line 29 and onward).

Hence we see the attitudes and dispositions of the advisor and the student interact closely with the task of seeking and giving advice as well as with the participants' role identities. Being a competent academic advisor means in part to be able to make clear to the students what is required, what is assumed, what is preferred, and what is permitted by the university. Complementarily, part of being a competent student advisee entails being able to project others' stances as well as their own so as to effectively elicit the advisor's advice.

Put more generally, the above analysis views the university institution as not merely a represented entity but a lived one. Institutions such as a university academic advising center do not just exist in the form of physical structure, personnel, and various rule books such as the university catalog, written policies regarding course credits, and so forth. They are lived by their members through seemingly routine actions, interactions, and activities. Knowledge and knowledgeability regarding institutional structures and constraints, institutional goals, and

institutional roles are produced and reproduced through the details of the participants' moment-by-moment conduct.

We have, to some extent, looked at the aspect of participants' conduct that is accomplished through language use. By language use, I have meant not only the use of vocabulary and grammar but also the sequential organization of talk. These two jointly provide the means through which institutions live and change. As academic counselors and student counselees talk to each other, as they ask and answer questions, as they tell and retell stories, as they quote others' speech, as they hedge or assert, they are actively engaged in reconstructing the institutional nature of their encounter.

3 Future Directions

Today, we witness a growing heterogeneity of national languages and cultures, increasing mobility of people across national, linguistic, and cultural borders, and global spread of information technologies (hypermedia and multimedia). Consequently, discourse analysts will not merely focus on a single language and a single culture, but will give greater attention to language and interaction in multilingual settings (Gardner and Wagner 2004; Nguyen and Kasper 2009; Zhu 2008; Li 2011) and intercultural and crosscultural communication (Kiesling and Paulston 2005). In addition to the traditional reliance on data transcripts, discourse analysts will explore the impact of new technology and investigate multimodality processes that involve visual, paralinguistic, and other semiotic resources (Bolden 2003; Garcia et al. 2009; Hutchby and Tanna 2008; Kress 2010). Motivated by a concern with the historical contingency and cultural continuity of language use, discourse analysts will also move beyond analyses of single snapshots of language interaction to speech events across time and space (Markee 2008; Wortham 2005; He 2011).

4 Discourse Analysis, Linguistics, and More

By way of concluding this chapter, let me say a few things to resituate our discussion both within and beyond linguistics. While it is correct to say that discourse analysis is a subfield of linguistics, it is also appropriate to say that discourse analysis goes beyond linguistics as it has been understood in the past. For as I have discussed above, discourse analysts research various aspects of language not as an end in themselves, but as a means to explore ways in which language forms are shaped by and shape the contexts of their use. Further, discourse analysis draws upon (and is practiced by scholars in) not only linguistics (especially functional linguistics) but also anthropology, sociology, psychology, philosophy, cognitive science, and other disciplines in the humanities and social sciences concerned with human communication.

Discourse analysis is a wonderfully creative enterprise. It is also a disciplined enterprise. It is creative in the sense that one can, for instance, combine interests in conversation analysis, grammar, storytelling, institutional discourse. and gender by investigating how gender is reflected and recreated through specific speech exchange systems and specific grammatical processes in conversational storytelling at workplaces. It is disciplined in the sense that not all approaches to discourse are equally defensible against all sources of doubt and that one needs to determine what constitutes the nature of the research question and to choose which set of theoretical and methodological constraints to abide by. Doing discourse analysis requires the researcher to closely engage with language and interaction through recordings, texts, or transcripts and to pursue the functions and meanings of language and interactional forms through *analyzing* (not merely displaying) them with analytic frameworks and procedures most suited for the research question.

Discourse analysis promotes a view of language which says that language use is not only reflective of other aspects of our lives but is also constitutive of them. In this sense, it revitalizes, advances, and systematizes functionally and anthropologically oriented schools of linguistics, thus creating a healthy balance with autonomist linguistics. As it draws insights from various disciplines, it also contributes to interfacing linguistics with other domains of inquiry, such that, for example, we might now investigate the construction of culture through conversation or program computers to generate interactive texts based on our understanding of the rules and principles of human interaction. Finally, discourse analysis brings to linguistics and related disciplines a human dimension. It focuses on language as it is used by real people with real intentions, emotions, and purposes. It reminds us that "language has a heart" (Ochs and Schieffelin 1989) and that language users and linguists do too.

REFERENCES

Antaki, C., Barnes, R., and Leudar, I. (2005). Diagnostic formulations in psychotherapy. *Discourse Studies*, *7*, 627–47.

Arminen, I., Auvinen, P., and Palukka, H. (2010) Repairs as the last orderly provided defense of safety in aviation. *Journal of Pragmatics*, 42(2), 443–65.

Atkinson, J. M. and Heritage, J. (eds) (1984). *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.

Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Oxford University Press.

Barnes, R. (2007). Formulations and the facilitation of common agreement in meetings talk. *Text & Talk*, 27(3), 273–96.

Barton, D. (2007). *Literacy: An Introduction to the Ecology of Written Language*. Oxford: Blackwell.

Block, D. (2007). *Second Language Identities*. New York: Continuum.

Blommaert, J. (2005). *Discourse: A Critical Introduction*. New York: Cambridge University Press.

Bucholtz, M. and Hall, K. (2005). Identity and interaction: a sociocultural linguistic approach. *Discourse Studies*, 7(4–5), 585–614.

Bolden, G. B. (2003) Multiple modalities in collaborative turn sequences. *Gesture*, 3, 187–212.

Capps, L. and Ochs, E. (1995). *Constructing Panic: The Discourse of Agoraphobia*. Cambridge, MA: Harvard University Press.

Cazden, C. (1988). *Classroom Discourse: The Language of Teaching and Learning*. Portsmouth, NH: Heinemann.

Chafe, W. (ed.) (1980) *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex.

Couper-Kuhlen, E. and Selting, M. (eds) (1996). *Prosody in Conversation*. Cambridge: Cambridge University Press.

Creese, A. (2008). Linguistic ethnography. In: K. A. King and N. H. Hornberger (eds), *Encyclopedia of Language and Education*, 2nd edn, Volume 10: Research Methods in Language and Education (pp. 229–41). New York: Springer.

Danes, F. (1974). Functional sentence perspective and the organization of the text. In F. Danes (ed.) *Papers on Functional Sentence Perspective* (pp. 106–28). The Hague: Mouton.

de Beaugrande, R. and Dressler, W. U. (1981). *Introduction to Text Linguistics*. London: Longman.

Drew, P. and Heritage, J. (eds) (1992). *Talk at Work*. Cambridge: Cambridge University Press.

Duranti, A. (1997). *Linguistic Anthropology*. New York: Cambridge University Press.

Duranti, A. (ed.) (2009). *Linguistic Anthropology: A Reader*, 2nd edn. Malden, MA: Wiley-Blackwell.

Duranti, A. and Goodwin, C. (eds) (1992). *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge: Cambridge University Press.

Duranti, A., Ochs, E., and Schieffelin, B. (eds) (2011). *The Handbook of Language Socialization*. Oxford: Blackwell.

Eggins, S. (2005). *An Introduction to Systemic Functional Linguistics*, 2nd edn. London: Bloomsbury Academic.

Ford, C. (1993). *Grammar in Interaction*. Cambridge: Cambridge University Press.

Ford, C. and Thompson, S. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, and S. A. Thompson (eds), *Interaction and Grammar* (pp. 134–84). Cambridge: Cambridge University Press.

Ford, C., Fox, B. A., and Thompson, S. A. (eds) (2002), *The Language of Turn and Sequence*. New York: Oxford University Press.

Fox, B. A. (1987). *Discourse Structure and Anaphora in Written and Conversational English*. Cambridge: Cambridge University Press.

Fox, B. A. (2007) Principles shaping grammatical practices: an exploration. *Discourse Studies*, 9, 299–318.

Fries, P. H. (1981). On the status of Theme: Arguments from discourse. *Forum Linguisticum*, 6(1), 1–38.

Fries, P. H. (2009). The textual metafunction as a site for a discussion of the goals of linguistics and techniques of linguistic analysis. In G. Thompson and G. Forey (eds), *Text Type and Texture* (pp. 8–44). London: Equinox.

Garcia, A. C., Standlee, A. I., Bechkoff, J., and Cui, Y. (2009). Ethnographic approaches to the internet and computer-mediated communication. *Journal of Contemporary Ethnography*, 38, 52–84.

Gardner, R. and Wagner, J. (eds) (2004). *Second Language Conversations*. London: Continuum.

Givon, T. (1983). *Topic Continuity in Discourse*. Amsterdam: John Benjamins.

Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96(3), 606–33.

Goodwin, C. (ed.) (2003) *Conversation and Brain Damage*. New York: Oxford University Press.

Goodwin, M. (2006). *The Hidden Life of Girls: Games of Stance, Status and Exclusion*. Oxford: Blackwell.

Gunnarsson, B-L., Linell, P., and Nordberg, B. (eds) (1997). *The Construction of Professional Discourse*. London and New York: Longman.

Halliday, M. A. K. (1978). *Language as a Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.

Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. London: Routledge.

Halliday, M. A. K. and Hasan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M. A. K. and Mathiessen, C. (2014). *Halliday's Introduction to Functional Grammar*, 4th edn. London: Routledge.

Hanks, W. (1990). *Referential Practice*. Chicago, IL: University of Chicago Press.

Hasan, R. and Fries, P. H. (eds) (1995). *On Subject and Theme*. Amsterdam: John Benjamins.

He, A. W. (1998). *Reconstructing Institutions: Language Use in Academic Counseling Encounters*. Greenwich, CT and London: Ablex.

He, A. W. (2011). Heritage language socialization. In A. Duranti, E. Ochs, and B. Schieffelin (eds), *The Handbook of Language Socialization* (pp. 587–609). Oxford: Blackwell.

Heath, S. B. (1983). *Ways with Words*. Cambridge: Cambridge University Press.

Heritage, J. (2005) Cognition in discourse. In H. te Molder and J. Potter (eds), *Conversation and Cognition* (pp. 184–202). Cambridge: Cambridge University Press.

Heritage, J. and Raymond, G. (2005). The terms of agreement: indexing epistemic authority and subordination in assessment sequences. *Social Psychology Quarterly*, 68, 15–38.

Hutchby, I. and Tanna, V. (2008). Aspects of sequential organization in text message exchange. *Discourse and Communication*, 2, 143–64.

Hutchby, I. and Wooffitt, R. (2008). *Conversation Analysis*, 2nd edn. Cambridge: Polity Press.

Hymes, D. (1974). *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia, PA: University of Pennsylvania Press.

Johnstone, B. (2008). *Discourse Analysis*. Oxford: Blackwell.

Kidwell, M. (2009) What happened?: an epistemics of before and after in "at-the-scene" police questioning preview. *Research on Language & Social Interaction*, 42(1), 20–41.

Kiesling, S. F. and Paulston, B.(eds) (2005). *Intercultural Discourse and Communication: The Essential Readings*. Oxford: Blackwell.

Koole, T. (2010). Displays of epistemic access: Student responses to teacher explanations. *Research on Language & Social Interaction*, 43(2), 183–209.

Kramsch, C. (ed.) (2002). *Language Acquisition and Language Socialization: Ecological Perspectives*. London: Continuum.

Kramsch, C. and Whiteside, A. (2008). Language ecology in multilingual settings: towards a theory of symbolic competence. *Applied Linguistics*, 29(4), 645–72.

Kress, G. (2010). *Multimodality: A Social Semiotic Approach to Contemporary Communication*. New York: Routledge.

Labov, W. (1972). The transformation of experience in narrative syntax. *Language in the inner city* (pp. 354-396). U. of Penn. Press.

Li, C. and Thompson, S. (1976). Subject and topic: a new typology of language. In C. N. Li (ed.) *Subject and Topic* (pp. 457–89). New York: Academic Press.

Li, W. (2011). Moment Analysis and translanguaging space: Discursive construction of identities by multilingual Chinese youth in Britain. *Journal of Pragmatics*, 43(5), 1222–35.

Mann, W. C. Mathiessen, C., and Thompson, S. (1992). Rhetorical structure theory and text analysis. In W. C. Mann and S. A. Thompson (eds), *Discourse Description: Diverse Linguistic Analyses of a Fund-raising Text* (pp. 39–78). Amsterdam: John Benjamins.

Markee, N. (2008). Toward a learning behavior tracking methodology for CA-for-SLA. *Applied Linguistics*, 29, 1–24.

Mathesius, V. (1975). On the information bearing structure of the sentence. In S. Kuno (ed.) *Harvard Studies in Syntax and Semantics* (pp. 467–80). Cambridge, MA: Harvard University Press.

Mendoza-Denton, N. (2007). *Homegirls: Language and Cultural Practices among Latina Youth Gangs*. Malden, MA: Blackwell.

Nguyen, H. and Kasper, G. (eds) (2009). *Talk-in-Interaction: Multilingual Perspectives*. Honolulu, Hawaii: National Foreign Language Resource Center.

Ochs, E. (1988). *Culture and Language Development*. Cambridge: Cambridge University Press.

Ochs, E. (1993). Constructing social identity: a language socialization perspective. *Research on Language and Social Interaction*, 26(3), 287–306.

Ochs, E., Schegloff, E. A., and Thompson, S. (eds) 1996. *Interaction and Grammar*. Cambridge: Cambridge University Press.

Ochs, E. and Schieffelin, B. (1989). Language has a heart. *Text*, 9(1), 7–25.

Peräkylä, A., Antaki, C., Vehviläinen, S., and Leudar, I. (eds) (2008). *Conversation Analysis and Psychotherapy*. Cambridge: Cambridge University Press.

Prince, E. (1981). Towards a taxonomy of given-new information. In P. Cole (ed.) *Radical Pragmatics* (pp. 223–56). New York: Academic Press.

Pütz, M., Neff-van Aertselaer, J., and van Dijk, T. A. (eds) (2004). *Communicating Ideologies: Multidisciplinary Perspectives on Language, Discourse and Social Practice*. Frankfurt/Main: Peter Lang.

Rampton, B. (1995). *Crossing: Language and Ethnicity among Adolescents*. London: Longman.

Rampton, B. (2006). *Language in Late Modernity: Interaction in an Urban School*. Cambridge: Cambridge University Press.

Sacks, H. (1984). On doing "being ordinary". In J. M. Atkinson and J. Heritage (eds), *Structures of Social Action: Studies in Conversation Analysis* (pp. 413–29). Cambridge: Cambridge University Press.

Sacks, H. (1992). *Lectures on Conversation*, vols I and II, edited by G. Jefferson, with an introduction by E. A. Schegloff. Cambridge: Blackwell.

Sacks, H., Schegloff, E.A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735.

Schank, R. C. and Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structure*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schaeffner, C. and Holmes, H. K. (eds) (1996). *Discourse as Ideologies*. Clevedon: Multilingual Matters.

Schegloff, E. A. (1979). The relevance of repair to a syntax-for-conversation. In T. Givon (ed.), *Syntax and Semantics, Vol.12: Discourse and Syntax* (pp. 261–86). New York: Academic Press.

Schegloff, E. A. (1996). Turn organization: one intersection of grammar and interaction. In E. Ochs, S. Thompson, and E. Schegloff (eds), *Interaction and Grammar* (pp. 52–133). Cambridge: Cambridge University Press.

Schegloff, E. A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis, Vol. 1*. New York: Cambridge University Press.

Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361–82.

Schegloff, E. A. and Sacks, H. (1973). Opening up closings. *Semiotica*, 8, 289–327.

Schieffelin, B. (1990). *The Give and Take of Everyday Life*. New York: Cambridge University Press.

Schieffelin, B. (2007). *Consequences of Contact: Language Ideology and Sociocultural Transformation in Pacific Societies*. New York: Oxford University Press.

Schiffrin, D. (1993). *Approaches to Discourse*. Oxford: Blackwell.

Schiffrin, D., Tannen, D., and Hamilton, H. (eds) (2003). *The Handbook of Discourse Analysis*. Wiley-Blackwell.

Scollon, R. and Scollon, S. (1981). *Narrative, Literacy and Face in Interethnic Communication*. Norwood, NJ: Ablex.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J. R. (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.

Seedhouse, P. (2004). *The Interactional Architecture of the Language Classroom: A Conversation Analysis Perspective*. Malden, MA: Blackwell.

Speer, S. A. and Stokoe, E. (eds) (2011). *Conversation and Gender*. Cambridge: Cambridge University Press.

Strauss, S. and Feiz, P. (2013). *Discourse Analysis: Putting our Worlds into Words*. London: Routledge.

Taboada, M. and Mann, W. C. (2006). Rhetorical Structure Theory: looking back and moving ahead. *Discourse Studies*, 8(3), 423–59.

Tannen, D., Kendall, S., and Gordon, C. (eds) (2007). *Family Talk: Discourse and Identity in Four American Families*. New York: Oxford University Press.

ten Have, P. (1999). *Doing Conversation Analysis*. London: Sage Publications

van Dijk, T. (ed.) (1997). *Discourse Studies* (2 vols). London: Sage Publications.

van Dijk, T. & Kintsch, W. (1983). *Strategies for Discourse Comprehension*. New York: Academic Press.

Van Lier, L. (2004). *The Ecology and Semiotics of Language Learning: A Sociocultural Perspective*. Dordrecht: Kluwer Academic.

Whalen, M. R. and Zimmerman, D. H. (1990). Describing trouble: practical epistemology in citizen calls to the police. *Language in Society*, 19, 465–92.

Widdowson, H. (2007). *Discourse Analysis*. New York: Oxford University Press.

Wierzbicka, A. (1992). *Semantics, Culture and Cognition*. New York: Oxford University Press.

Wilce, J. (2009). *Language and Emotion*. New York: Cambridge University Press.

Wodak, R. and Meyer, M. (2009). *Methods of Critical Discourse Analysis*, 2nd edn. London: Sage Publications.

Wortham, S. (2005). Socialization beyond the speech event. *Journal of Linguistic Anthropology*, 15(1), 95–112.

Young, R. (2009). *Discursive Practice in Language Learning and Teaching*. Malden, MA: Wiley-Blackwell.

Young, R. and He, A. W. (eds) (1998). *Talking and Testing: Discourse Approaches to the Assessment of Oral Language Proficiency*. Amsterdam: John Benjamins.

Zhu, H. (2008). Dueling languages, dueling values: code-switching in bilingual intergenerational conflict talk in diasporic families. *Journal of Pragmatics*, 40, 1799–1816.

22 Linguistics and Literature*

NIGEL FABB

1 Literary Language and Its Distinctive Characteristics

"Literature" is a name for an artwork which is made from language, and literary texts can be called "verbal art." Literary texts have superadded regularities, which are in addition to the regularities of ordinary language. For example, an English sonnet is in 14 lines: neither the division into units which are lines, nor the counting of any kind of unit, are a type of rule or form of ordinary language, and so these are examples of superadded form. The kinds of form may be variants of ordinary linguistic form: for example, there are clear structural and articulatory/phonological similarities in principle between prosodic phrases and poetic lines. Furthermore, the superadded form is sensitive to the forms of ordinary language; for example, line-ends usually coincide with word-ends, and metrical form is based on linguistic units (syllables and morae). In an English sonnet, the line has 10 syllables, and this both draws on linguistic form (the syllable) and does something nonlinguistic with it (counts up to ten). Note that this is a fact about the text, not about how it is laid out on the page or spoken: even if laid out as prose, it is a fact that every tenth syllable is word-final, and this is because lineation (in 10-syllable lines which must end at word ends) is an abstract fact about the text, not about how it is performed or displayed. The sonnet's syllables are in a characteristic rhythm relative to the line, such that even-numbered syllables (counting anew from the beginning of each line) tend to be stressed, as can be heard in the following four lines from Shakespeare's sonnet 82.

> I grant thou wert not married to my Muse,
> And therefore mayst without attaint o'erlook
> The dedicated words which writers use
> Of their fair subject, blessing every book.

Linguists have argued that the control over rhythm is by rules or conditions of the kind that regulate stress in ordinary words, so, again, the literary form adapts linguistic form. The lines also rhyme, and this both depends on linguistic form – the structure of the syllable and the relations of identity between sounds – and does something nonlinguistic with it – organizes the rhyme into an alternating ABAB pattern. In sum, a literary text has form which is superadded; the

* For comments on the draft of this article, I am grateful to the editors and Gary Thoms.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

superadded form refers to linguistic form (sometimes in complex ways), and may itself be organized in ways similar to linguistic form (or be a version of linguistic form), while also being subject to counting and patterns which are quite distinct from what we find in ordinary language.

The study of linguistics and literature is an attempt to tease apart the various ways in which the superadded forms of literature exploit ordinary linguistic form (or do something nonlinguistic with it). Fabb (2010) characterizes as "the development hypothesis" the view that the literary forms of a language are developments of the linguistic forms of that language. The looser version of this hypothesis allows a literature to develop a form from a type of linguistic form which may not even appear in its source languages; thus Yip (1999) argues that sound patterns in English may be manifestations of the linguistic process of reduplication, a type of linguistic form which does not appear in English. The notion of stricter dependence of a literature on its specific language is expressed by Sapir, who says

> study carefully the phonetic system of a language, above all its dynamic features [i.e., prosody], and you can tell what kind of a verse it has developed – or, if history has played pranks with its psychology, what kind of verse it should have developed and some day will. (Sapir 1967: 230)

A seminal claim of the grounding of literary form in the linguistic forms of language is expressed by Kiparsky (1973: 235): "The linguistic sames which are potentially relevant in poetry are just those which are potentially relevant in grammar." Kiparsky notes that poetry at all times and in all places is similar within a limited range of variation; in this, poetry differs from other arts (e.g., the visual arts, which he says vary much more greatly), and this is because poetry has direct and privileged access to universals of human cognition as expressed in universals of language.

Literary linguistics is almost always a study of texts or collections of texts, for example, all the iambic pentameter lines written by Shakespeare, or a collection of folksongs. It is rare for literary linguists to directly study the judgements of speakers (or original audience). Some of the literature-specific regularities which can be discovered in any collection or corpus are subject to variation. For example, a corpus consisting of lines in English iambic pentameter is likely to have the 10-syllable line as most common, but also some 11-syllable lines, and in some cases longer or shorter lines as well. The same corpus will have some lines with fully periodic rhythms (every odd numbered syllable having less stress than the adjacent even-numbered syllables), but there will be many other rhythmic patterns. These regularities and variations can be represented statistically, as descriptions of the corpus, and this has been a characteristic of Russian literary linguistic approaches (e.g., Tarlinskaja 1976). The statistical regularities can be built into the system of poetic rules: This is the goal of maxent grammars for meter (e.g., Hayes and Moore-Cantwell 2011) in which the metrical grammar consists of a set of weighted constraints which fully describe all the variations in the corpus.

Approaches which seek to capture a whole corpus in a single grammar are faced with the difficulty that there are multiple causal factors impinging on a literary text which can lead to variation in a corpus. The Russian Formalists argued that variation or deviation is the fundamental defining characteristic of the literary, so that "deviant" novels such as Lawrence Sterne's *Tristram Shandy* become exemplars of literature rather than outliers. While formal deviance as a defining characteristic cannot be true of all literature, it shows that variation is not just a statistical property of a corpus relative to a regularity, but can have many functions. For example, literary texts show localized variation (they have marked form) in specific places of major functional importance in the text as a whole – at the beginning or the end, at an episode boundary (Hymes 1981), and at a key narrative point (Labov and Waletzky 1967).

There may be less specifically to be said about literature when it comes to the semantics and pragmatics of a literary text. Consider for example a Relevance Theoretic perspective (Sperber and Wilson 1995), according to which an author communicates an informative intention (a set

of thoughts) by producing a text; the text constitutes partial evidence for the informative intention. This partial evidence comes for example from the logical forms which can be derived by syntactic/semantic rules. Such logical forms could be derived from the language of literary texts just as they are in ordinary texts, and the pragmatic processes should be just the same. It is difficult to see how literary texts could communicate meaning in a manner which is categorially different from other uses of language such that literature would need an adapted semantics or pragmatics (analogous to the "adapted phonology" proposed for metrical rules, and perhaps an "adapted syntax"). In some cases, there are imposed rules of interpretation, but these are laid over ordinary semantic/pragmatic processes rather than intervening in them. For example, in (Papuan) Asmat songs, words can be systematically reinterpreted as meaning something other than what they encode; the word *cem* means "carrying bag" but in a poem can be interpreted to mean "house" (Voorhoeve 1977). This is an imposed recoding of the word, and analogous to the nonlinguistic manipulation of language: a nonsemantic/pragmatic manipulation of meaning. Similarly, meaning might be coded in literary texts in ways which do not relate to language at all, as in pattern poems where the text is shaped on the page to form an image.

2 Poetry: Text Divided into Lines

In many (perhaps all) literatures we find texts which are divided into a sequence of sections, called lines, and are called "poems" (or "verse," a term sometimes used to characterize the specifically formal aspects of poetry). In many traditions the poems are sung. The line is a type of section which is not determined by syntactic or prosodic constituency, though the line may coincide with syntactic or prosodic constituents. There is no single defining feature of a line such that we can guarantee that a line in some tradition will necessarily have that feature; instead we can identify features which are true of lines in different subsets of the traditions, in different combinations. These characteristics all raise the question of what lines are, whether they have any specific cognitive status, and whether they have some common function, which is returned to at the end of this section.

I now list some of the characteristic features of lines across a range of traditions. For each feature, it is always possible to find a type of line in some tradition which lacks that feature, which means that lines have no necessary characteristics.

(i) The line ends with a visual break (on the page).
(ii) The line ends on a pause (in speech), and more generally may end on a prosodic phrase boundary.
(iii) The final syllable of the line is lengthened (in performance), another form associating the end of the line with the end of the prosodic phrase.
(iv) Again, like prosodic phrases, there may be a characteristic pitch change over the final syllables of the line (e.g., pitch always falls).
(v) There is a characteristic line-final word or vocable (a nonword, often a single vowel).
(vi) The end of the line coincides with the end of a word (though this is almost always true, there are some traditions in which it is not a requirement, such as Greek lyric poetry, where a word can be split by a line boundary).
(vii) The end of the line coincides with a syntactic boundary (e.g., the end of a phrase, of a sentence).
(viii) The line is semantically independent (a "sense unit").
(ix) The line is divided into sublines such as half-lines.
(x) The line is combined into super-lines, such as a couplet.
(xi) The line is metrical.

(xii) Where the line has a regular rhythm, the rhythm tends to be looser toward the beginning of the line and stricter toward the end, and sometimes only the end of the line is metrical.

(xiii) In a quantitative meter, a light syllable can be used in line-final position where a heavy syllable would be expected (called "brevis in longo," and perhaps related to the lengthening of final syllables in some performed traditions, which may be a general prosodic phenomenon).

(xiv) There may be a controlled asymmetry: for example, in lines in the Finnish Kalevala longer words come later in the line, while, in Kaytete songs, longer words come earlier in the line (Turpin 2007).

(xv) Rhyme or alliteration is located relative to the line (e.g., line-final or line-initial, or there may be some requirement that the line contains one rhyming or one alliterating syllable, etc.).

(xvi) The line is a counted unit (e.g., quatrains, sonnets, etc. all count lines).

(xvii) The line has a systematic relation to a musical tune (a "text-to-tune matching" may be line sensitive).

(xviii) There is syntactic/lexical parallelism between two halves of the line or between the line and an adjacent line, usually manifested by similar syntactic structures with different but related lexical items.

One of the interesting questions about poetry relates to singing: Much poetry is sung poetry, and can be set to a rhythmically regular music. In this case, we might ask what aspects of the linguistic form of the text are determined by the music to which it is set, or what complex relations hold between the independent forms of music and language. These issues are discussed, for example, by Dell and Halle (2009) and Dell and Elmedlaoui (2008).

Poetry is divided into sections on various levels: Parts of lines (hemistichs) or groups of lines (couplets, stanzas, etc) are also kinds of section. Hayes and MacEachern (1996) suggest that, in some types of poetry, the line is the level of sectioning at which prosodic and poetic sections show the greatest coincidence. What are lines for? One line of enquiry might look at cognition or performance, and whether the line will fit into working memory (Fabb 2014). Current accounts of working memory (Baddeley 2012) suggest that up to about 15 syntactically connected words can be held at the same time in working memory; this is easily enough to accommodate almost all poetic lines, along with other material. This makes psychological sense of Jakobson's (1960) proposal that the poetic function (often associated with text divided into lines) is performed when attention is drawn to the verbatim form of the text (Jakobson's focus on the "message"); verbatim form exists in full only in working memory, after which it takes on more schematic forms associated with other memory and processing systems (see also Tsur 1998). In Fabb (2015), I argue that "the line" can be understood as both the section to which all the added forms, of meter, rhyme, alliteration, and canonic parallelism refer and also the section which is sufficiently small to be held as a whole verbatim sequence in working memory.

3 Metrical Poetry

The kind of poetry which has drawn most attention from linguists is metrical poetry, and in particular the kind of metrical poetry which has a predictable rhythm (as opposed to metrical poetry in which morae or syllables are counted without any determinate rhythm). The rhythm is predictable both in the sense that the line as a whole has a rhythmic profile, and in many cases there are also rhythmic units (feet, Greek *metron*, Arabic *ajzāʔ*, Sanskrit *gaṇa*, etc.) within the line which repeat to produce a periodic rhythm. In rhythmic metrical poetry there are two

constraints on the line: the line is of a certain length (counting syllables or morae) within a permitted range of variation, and there is a predictable rhythm based on stress, syllable weight, or tone. I now give examples of some types of meter, beginning with English iambic pentameter.

```
(1) Of man's first disobedience and the fruit
    1   2     3     4 5 6 *7    8     9    10
        s     s         s                s       stress
    Of that forbidden tree whose mortal taste
    1   2   3  4   5     6    7   8   9   10
        s       s       s        s       s        stress
```

The lines in (1) are the first two lines of Milton's *Paradise Lost* and are in "iambic pentameter" which normatively has 10 syllables in each line, and even-numbered syllables are more stressed than adjacent odd-numbered syllables (a pattern seen throughout the second line, where *s* indicates a syllable likely to have greater stress). In some accounts iambic pentameter is treated as a pattern of five repeated binary feet (i.e., pentameter), with the second part of the foot stronger than the first (i.e., iambic). Note that the first line has a different rhythm while loosely conforming to this pattern, and this is permitted by this meter. The number of syllables can also vary, within specific limits; for example, in "-ience" what might be two syllables are counted as one for metrical purposes because a vowel precedes a vowel.

The Latin lines in (2) are the first two lines of Virgil's *Aeneid* and are in "dactylic hexameter."

```
(2)   Arma virumque cano, Troiae qui primus ab  oris
       1  2  3 4   5  6 7    8 9    10 11 12 13 14 15    syllables
       h  l  l h   l  l h    h h    h   h l  l  h  h     weights
      | 1     | 2      | 3    | 4      | 5     | 6   |    feet

       Italiam fato profugus Lavinaque    venit
       1 2 3 4   5 6   7 8 9  10 11 12 13  14 15    syllables
       h l l h   h h   l l h  h  h  l  l   h  h     weights
      | 1   | 2   | 3     | 4   | 5      | 6    |    feet
```

This is a quantitative meter in which the rhythm is based on syllable weight (a pattern of heavy "h" and light "l" syllables) rather than stress. Thus in the word "arma," "ar" is a heavy syllable because it ends on a consonant (it would also be heavy if it contained a long vowel) and "ma" is a light syllable (because it ends on a short vowel). The meter controls where light and heavy syllables can fall. The syllables are combined into six units ("feet") which are either heavy-light-light (hll) or heavy-heavy (hh) in syllable weights, in different patterns. (Though these two lines both have 15 syllables they could have between 13 and 17 syllables.) This meter also has a bridge rule which stipulates that there must not be a word boundary between the third and fourth feet (in these lines "Troiae" and "profugus" are spread across these feet).

The French lines in (3) are the first two lines of Jean Racine's *Phèdre* and are in the "alexandrine" meter, which normatively has 12 syllables, and has no characteristic rhythm other than that it requires the sixth and twelfth syllables to be stressed and word-final (a "caesura rule").

```
(3)  Le dessein en est pris : je pars, cher Théramène,
     1   2  3   4  5     6     7   8     9  10 11 12

     Et quitte le séjour : de l'aimable Trézène.
     1    2  3  4  5  6     7   8  9  10 11 12
```

This is an example of a meter in which there is counting but (in most analyses) no general control of rhythm.

In addition to the metrical types exemplified by (1)–(3), there are also meters where the rhythmic rules are sensitive to lexical tone (Vietnamese, Chinese); moraic meters where the rules count morae (i.e., a heavy syllable counts as two morae, a light as one mora) such as Japanese; and meters which count morae and also control the distribution of syllables (Hindi, Berber, Somali); along with meters which control both quantity and stress (Finnish), and other types. The classification and analysis of some meters is subject to debate; thus, for example, while standard analyses of Old English meter treat it as a meter based on patterns of stress, Golston and Riad (2003) argue that it is instead based on patterns of quantity.

Meters partition syllables into two kinds, usually based on stress, or quantity, or tone; thus Lotz (1960: 140) says that "it is interesting to note that the phonological elements are grouped into two base classes, never into more." For example, in a language which has several levels of stress (as English does), or has light, heavy, and superheavy syllables, the meters tend nevertheless to distinguish syllables into two classes. A striking example is presented by Vietnamese, which has six phonemic tones, but only two tonal types of syllable are differentiated in determining the pattern required for the line of poetry. Again we see that verbal art is based on linguistic form, but does something additional (perhaps "nonlinguistic") to it.

Poetic meter has elicited the most complex linguistic accounts of literary form, and I now consider four approaches. Hanson and Kiparsky (1996) provide one approach, in which a meter is a template consisting of left- or right-headed binary feet, each consisting of a strong metrical position (its head) and a weak metrical position. For each metrical position, a parameter is set which matches it with a phonological constituent up to a certain size, where the upper limit can be: a mora, syllable, phonological foot, or phonological word. Weak or strong positions are restricted relative to nonprominent or prominent phonological constituents. The phonetic correlates of "prominence" involve the setting of a parameter defined by syllable weight, stress, strength (a subset of stressed syllables are strong), or pitch accent, or some combination of these. As an example of how a meter sets parameters, consider English iambic pentameter: A phonological foot (of one or two syllables) can match a metrical position, weak positions must not be filled with prominent phonological constituents, and prominence is defined in terms of syllable strength. Thus some stressed syllables are forbidden from appearing in weak metrical positions, and up to two syllables can be matched to a single metrical position. In this way, rhythm and syllable count are both controlled by the meter. Each language has metrical forms which are largely derived from the phonology of the language. A meter has its parameters set such that it tends to enable the most flexible use of the vocabulary of the language: that is, the meter "fits" the language. The parameters also maximize the aesthetic interest of the poetry, by, for example, warding off monotony. This negotiation between fit and interest is the aesthetic basis of meter in their account. Hanson and Kiparsky's article develops from earlier work by each of the authors (e.g., Kiparsky 1977; Hanson 1992), which is characterized by as precise as possible accounts of the metrical conditions operating for a specific poet (and tradition).

In a second approach to poetic meter, Hayes and MacEachern's (1998) stochastic Optimality Theory incorporates the statistical properties of variation over a corpus into the rules generating the meter of that corpus. The focus of this article is on the well-formedness judgements for four-line units (quatrains) in English folk poetry. Lines are matched to templates (grids) which have four strong positions, and there are different types of "cadence," depending on what happens after the third strong position in the grid. The most cadential ending has the third strong position filled by a stressed syllable which is line-final, so that the fourth position is empty. The next most cadential ending has an unstressed syllable following this. The next has the third and fourth strong positions filled by adjacent stressed syllables. The least cadential ending has the fourth strong position filled and separated by an unstressed syllable from the third strong position. A group of lines forms a salient metrical constituent if the lines have the same type of cadence but

the final line is more cadential than the others. Though more cadential lines have the advantage of emphasizing group structure, they are rhythmically less structured as lines. Thus there is no "winning" combination of line types: Different types of quatrain resolve the conflicting aesthetic demands of the line and the group in different ways. This account is framed by Optimality Theory in terms of freely rankable violable constraints; different possible rankings derive the types of quatrains which are preferred in the corpus. The greater frequency of some types of quatrain over others can be modeled by attaching strictness ranges to constraints: though in principle the constraints can be re-ranked in order to generate all the various types of extant corpus, the re-ranking itself is controlled by preference rules.

In another approach, Golston and Riad (2000) take an Optimality Theoretic approach, according to which the basic principle is that different kinds of meter show different kinds of markedness, respecting or violating constraints in different ways (often violating "rhythmic constraints" and so a meter is not the same as rhythm). The basic constituent is taken to be the actual stress foot in the phonology of the language in question: In Greek and Arabic this is the moraic trochee (Hayes 1995: 125), which is a mora–mora pair in which the first is more prominent than the second (a light syllable is a single mora, a heavy syllable two morae). In Greek epic and spoken poetry there are various kinds of foot, which are either unmarked or marked. If feet are in the unmarked case binary, then a Greek anapaestic foot is the most unmarked foot because it consists of two pairs of morae, realized as HLL, LLH, HH, or LLLL (all possible variants of the anapaestic foot). Greek dactylic feet are realized as HH or HLL which means that they are unmarked for binarity (they consist of four morae) but are marked for stress clash, because HH involves adjacent stressed syllables (H is stressed), and so does HLL (because the first L in a pair is stressed). Different lengths of line (measured in feet) also involve specific violations of specific constraints, which group feet into pairs (metra), metra into pairs (lines), etc. A line with no violations of binarity would have four feet, with other lengths of line violating binarity on one or more levels of structure.

In a fourth approach, Fabb and Halle (2008) begin from the line as the metrically relevant unit. Hanson and Kiparsky apply rules or conditions to the relation between a phonological unit (the syllable, or the phonological foot, or the mora) and a metrical unit (the metrical position, usually differentiated as W or S): metricality is thus typically a local relation between a few syllables and a small metrical structure. In contrast, Fabb and Halle subordinate all other metrical considerations to the fact that the sequence of syllables forms a line: thus any line-internal syllable is controlled by metrical rules or conditions only by virtue of its location relative to the line as a whole. In this theory, the whole line is taken as input to a set of rules which generate a grid, step by step, by projecting asterisks and grouping them by inserting parentheses; asterisks project from syllables in the first instance, and then each group projects one asterisk to the next gridline, so that the gridlines contain fewer asterisks until the final gridline has just one. The final output of the grid-building rules for iambic pentameter, when applied to the first line of *Paradise Lost*, is shown below.

```
(4)  Of man's first disobedience and the fruit
    )*    *)     *      *) * *) *     *)     *    *)        gridline 0
          *             *)   *        *          *)        gridline 1
                       (*                       *(        gridline 2
                        *                                  gridline 3
```

In this approach, the fundamental aspect of the meter is a set of rules which generates a well-formed grid only from a line of the right length (the iambic pentameter rules generate a grid from a line of 10 syllables). The rhythm or other patterns in the line are secondary, and are derived from the grid. So, in the case of iambic pentameter, a condition requires that certain stressed syllables (maxima) must project to gridline 1; there is only one such syllable in this line, the third syllable

of "disobEdience" and this does project to gridline 1 so the condition is satisfied. Note that the grid does not directly represent the rhythm, and has characteristics which are not reflected in the phonology of the line; of all the theories of meter this is the least based in phonology, and the one which most proposes that meter is a different system from phonology. By referring to the grid which projects from a line, it is possible for conditions to identify specific syllables within the line as having specific characteristics, including rhythmic characteristics, but also as being word-final or rhyming, or some other characteristic. The rules for each meter will only generate a well-formed grid from a line with a specific number of syllables, and so the metrical rules "count syllables"; the counting mechanism produces as its output the grid which can then have other conditions attached to it. This approach has a wider typological range than most of the other theories, because it covers nonrhythmic meters and aperiodic rhythmic meters. However, this theory is also less precise in its control of exactly which types of line are and are not possible in a particular meter. Fabb (2002) addresses the underdetermining of rhythm in the Fabb and Halle approach, and suggests that rhythmic regularities may also arise from the operation of processes which are not part of the metrical generative rules such as the need to communicate to a hearer various formal properties of the text.

4 Sound Patterning in Poetry

A common kind of regularity in poetry involves repetition of sounds or sequences of sounds. Most such repetitions are sensitive to the structure of the syllable, that is, repetitions which exclude a syllable onset are called "rhyme" and other repetitions are called "alliteration."

Linguistic research in this area has focused primarily on cases where a pair of sound-sequences (e.g., in rhyme) are treated for poetic purposes as the same, but are not phonetically identical. For example, many kinds of rhyme may accept two syllables as rhyming where the vowels are identical, but the final consonants differ (e.g., Zwicky 1976). Similarly, it is common to find that vowel-initial words are counted as alliterating even though the vowels may be different. Most accounts of these phenomena seek to explain the pairing of dissimilar sounds in linguistic terms. One approach is to argue that similarity of underlying phonological representations can be recognized by sound patterning rules, even if surface representations are dissimilar (e.g., Malone 1988). Other approaches use Optimality Theory in explaining the similarity between sounds (Holtman 1996; Steriade 2003). Steriade looks at half-rhymes in Romanian poetry, and shows that poets admit only certain types of half-rhymes and systematically prefer certain half-rhymes over others. The half-rhymes all resemble familiar phonological processes, but not processes which are active in Romanian: For example, a common half-rhyme involves voiced vs. voiceless consonants after a nasal, which resembles the phonological process of postnasal voicing but this process is not found in Romanian. Steriade argues that half-rhymes are accepted (and preferred) based on knowledge of the relative perceptibility of specific sounds in specific contexts: in certain contexts, the cues differentiating two sounds are less salient and so a half-rhyme is more possible. Tristram (1995) argues that near-sameness rather than identity is preferred in Irish, Welsh, and Old English metrical poetry, including rhyme and alliteration, perhaps for aesthetic reasons. Kawahara (2007) explores the role of near-similarity in both language and poetry.

Another aspect of sound patterning is its distribution in a text: Most kinds of sound patterning are sensitive to the line as a structure. For example, the sound patterning may hold within the line, or between two lines, or between the end of one line and the middle of the next. This again raises the possibility that the function of the line is to provide a structure or frame upon with another kind of regularity (rhyme, alliteration, etc.) can be mounted. Note, for example, that while rhymes can be located in various places relative to a line (usually at the end but sometimes midline), a pair of rhyming words is never located just relative to each other (e.g., six syllables apart). Fabb (1999) argues, on the basis of a limited typological survey, that while rhyme can appear in

ABAB patterns, this is not possible for systematic alliteration (e.g., line-initial alliteration involves adjacent lines). If correct, this would show both that alliteration and rhyme are different in kind (rather than two variants of essentially the same process), and also that they are sensitive to the domains within and across which they operate. Rhyme and alliteration might be considered composite forms: In each case, they involve a pairing of words based on phonological similarities (one kind of form) and the distribution of those words relative to lines in a complex pattern (a different kind of form); the two kinds of form do not generally interact, so there is rarely any dependence of pattern on phonology.

An explanation of the aesthetic function of sound patterning comes from pragmatics; Pilkington (2000) examines various aspects of literary language, in particular stylistic features of poetry such as sound patterning, and using Relevance Theory presents an explanation of why they give rise to an aesthetic effect. He suggests that stylistic features have an effect on our processing of concepts associated with the words of the poem. The stylistic features demand additional computations which slow down the processing of the poem, and so give the hearer more time to access the concepts associated with the words in the line, and hence more opportunity to access the assumptions stored in the encyclopaedic entries for these concepts. By this means, the additional computations demanded by the stylistic effects give rise to a larger and more weakly implicated set of assumptions, which Sperber and Wilson (1995: 222) describe as "the peculiar effect of an utterance which achieves most of its relevance through a wide array of weak implicatures"; they call this a "poetic effect." In keeping with the notion that aesthetics is a way of knowing (i.e., has an epistemic component), it is possible that the experience of this wide array is itself one of the kinds of aesthetic experience. MacMahon (2007) further pursues these notions, taking a psycholinguistic approach to the ways in which literary form forces a type of processing which produces a variety of specifically literary effects.

5 Parallelism

There are three major kinds of superadded form in poetry: meter, sound patterning, and parallelism. Parallelism is a relation between sections of text which are similar in some linguistic structural manner, but differ in specific ways. For example, a common pattern in Eastern Indonesian poetry is to have two lines which have the same syntactic structure but differ in a specific pair of lexical items, one in each line in specific places; they may be commonly associated as a pair, and may have similar or different meanings (Fox 1988). The pairing of items without much associated syntactic similarity may be more dominant in other traditions (such as Hebrew). In some traditions, parallelism is emergent, and can involve complex hierarchical structures, as shown for Nahuatl by Bright (1990). One of the more striking characteristics of many parallelistic traditions is where a pair of words exists as a pre-existing resource; thus there are fixed "dyads" in Indonesian poetry (Forth 1988), and in Nahuatl pairs of words in parallel structure produce a combined idiomatic meaning which is not derived from the component parts (called "difrasismo," Léon-Portilla 1985). Parallelism is an important formal device in many literatures, which has been extensively described, and its ethnopoetic properties examined; it has not yet been subjected to the same level of theoretical linguistic scrutiny as has been applied to meter and sound patterning.

6 The Syntax of Poetry

A line of poetry is not itself a linguistic constituent, and need not be coextensive with a linguistic constituent (i.e., sometimes verse-lines are whole sentences, but they need not be). The fact that syntactic constituents in particular can be split across verse-line boundaries ("enjambment") is

well known; in some traditions, even words can be split across verse-lines. Metrical and other formal requirements on the verse-line may require certain syntactic rearrangements or deletions which are not available in ordinary language, and in some cases this syntactic freedom seems to be welcomed irrespective of whether the meter forces it (e.g., Icelandic, discussed below). Syntax in poetry includes reorderings without any clear syntactic or functional motive (such as information structuring); reorderings which extract constituents out of phrases which are normally immune to having material extracted ("islands"); and ellipses/gapping of nonconstituents. Austin (1984) offers a standard discussion of these phenomena; one of his examples of reorderings which either have no clear functional motive or which is ungrammatical is from Shelley's "Adonais": "Not all to that bright station dared to climb; / And happier they their happiness who knew."

The unusual syntax of poetry was addressed by Levin (1964), who proposes that the grammar be altered so that it is capable of generating also these poetic texts. In response, Thorne (1965) suggests that a single poem might be considered as in a language of its own, and hence that a grammar be formed specifically to generate this poem. Hendricks (1969) offers a useful discussion of this debate. Austin (1984) suggests that the existing grammar can sufficiently explain poetic language, if we see specific lines as violating specific rules (e.g., the complex noun phrase constraint might be violated in a certain line). In Icelandic *dróttkvaett* poems a noun phrase can have the noun moved to one end of the clause and its determiner to the other; O'Neil (2001) argues that though the movement involves a "relaxation" of ordinary syntactic rules and borrowing of rules from other languages (e.g., Japanese-style scrambling), items are nevertheless moved to syntactically defined landing sites.

The specific relation between metrical and syntactic rules was discussed by Fowler (1966), and is addressed by Youmans (1983) who looks at reorderings which are forced by the meter as a way of comparing metrical theories. Hayes (2003 [1989]) argues that, in metrical lines, the rules of the meter take no account of syntactic structure, but instead are dependent on prosodic phrasing (which in turn is based on syntactic structure); again, syntax is de-emphasized. In Optimality Theoretic approaches, it has been argued that syntactic constraints and prosodic constraints are ranked relative to one another, with poetry-specific grammars (poetic grammars) allowing prosodic constraints to outrank syntactic constraints. Thus Golston and Riad (2000: 103) suggest that "in poetry, prosody outranks syntax"; Fitzgerald (2007) argues that some syntactic rules outrank some prosodic rules and are outranked by others.

7 The Component Parts of a Narrative

Labov and Waletzky (1967), based on their corpus of contemporary urban oral narratives, is one of the most important sources for the linguistic study of narrative (see Bamberg 1997). The authors argue that the main body of the narrative consists of a complication followed by a resolution. The main body may be preceded by an orientation, and followed by a coda. The shift to resolution may be signalled in the narrative by various formal means, including repetition, tense shifts, use of modals, etc. This signalling is part of the general practice of "evaluation" of the narrative by which the narrative draws the hearer's attention to the point of the narrative, by evaluating the significance of events reported.

Just as the shift from complication to resolution is linguistically marked, so the boundaries between episodes (sections with differentiated content) are marked in a narrative. For example, in an oral narrative the conjunctions "so" and "and" may organize the clauses they introduce into different kinds of units (e.g., "so" may be reserved to mark a new episode; "and" may be used at the beginning of a small group of clauses with a particular structural role). Fabb (1997) summarizes research showing that a range of different linguistic forms can be used to mark episode boundaries: connectives, hearsay particles, shifts in tense, repetition, and parallelism, for example. Hymes (1981) argues that episodes are also part of a counting system which can

pervade a narrative (depending on the tradition) which he calls a "pattern number." Two common types are a 3–5 pattern, and a 2–4 pattern. For example, in a 3–5 pattern, episodes might be divided into three parts (three clauses) and five episodes might make up the story (or vice versa); furthermore, narrative contents might repeat the pattern with characters or objects or some other aspect of content – thus, a heroine might be one of three sisters and have to walk five miles to undertake a task.

Labov and Waletzky (1967) distinguish between clauses which constitute the story proper, and clauses which give background information; the former can be defined as clauses whose linear order is the same as the linear order of the events. The distinction between what are generally called "storyline clauses" and "nonstoryline clauses" has been investigated for many literatures; storyline clauses often have distinctive formal characteristics, which may for example include giving greater prominence to the verb (by placing it at the edge of the clause). Hopper and Thompson (1980) show that storyline clauses are typically higher in transitivity features, based on an extended notion of transitivity (e.g., a verb with an individuated object is measured as more transitive than a verb with a mass object).

8 The Representation of Thought and Speech

Literary texts represent actions and events and states of affairs, often in a fictional world. These representations are almost always best understood as metarepresentations, often recursively embedded (i.e., meta-metarepresentations, etc.). A metarepresentation is a representation of a representation, and is a possibility in all representational media: A photograph of a painting and a photograph of a photograph are both examples of metarepresentations. Language is particularly well suited to metarepresentation, because of its recursive character, which enables a sentence to contain a sentence. Sperber and Wilson (1985/6) argue that irony can be understood in terms of metarepresentation: Irony is a type of "echoic utterance," which is a mode of metarepresentation in which the framing sentence indicates the attitude of its producer to the framed sentence.

The term "metarepresentation" has gained wider currency particularly since Sperber (2000), but the problems of metarepresentation in literature have been extensively discussed, often under other names. For example, Banfield (1982) looked at some modes of metarepresentation in nineteenth-century fiction which she argued are possible only in writing (hence "unspeakable sentences," the title of her book). Similar concerns are dealt with under the heading of "focalization" in narratology (representations are focalized through a subject, i.e., metarepresented). There is a standard critical terminology based on a typology of kinds of metarepresentation, which includes terms such as "direct speech," "direct thought," "free indirect speech" (or *erlebte Rede* or *style indirect libre*), and so on. This terminology differentiates instances of metarepresentation based on what is being metarepresented (speech or thought), and how closely the framed sentence resembles what it represents; in "indirect" types, the framed sentence might have a different tense from the sentence which it represents, for example.

Work on metarepresentation is usually incorporated into the broader theory that humans have a specialized capacity for "theory of mind," which is our knowledge that humans have minds and our ability to attribute thoughts to ourselves and others (to mind-read). Zunshine (2006) discusses how our theory of mind capacity is exploited in fiction, where we are asked to make complex assessments of the thoughts of the author and characters, including characters' thoughts about other characters' thoughts. Zunshine raises the question of how many levels of recursion are found in different types of literature, and whether this tells us anything about cognitive limits on metarepresentation. Zunshine also discusses how different modes of metarepresentation are tied to the technology or medium of a text, such as whether it is in writing or in speech.

In a reworking of Labov and Waletzky (1967), Fludernik (1996, 2003) develops what she calls a "natural narratology," which looks at different modes of narrativization as involving the expression of experience (e.g., the narrator's experience) through different perspectives. Fludernik argues that the perspectives are drawn from five cognitive frames which are generally available, such as our general familiarity with "telling" (a story) or being a witness to events, or our ability to be immersed in an experience. Different types of narrative are explained by their emphasis on different cognitive frames; for example, narratives with no apparent narrator emphasize the "viewing" frame (but for example lack "experiencing" as a frame). Fludernik's goal is to explain the human reworking of experience in terms of its emotional and evaluative significance. Fludernik's approach is one of the many current approaches to literary linguistics which seek to explain linguistic phenomena (here, narrative and the representation of experience) as dependent on broad cognitive principles; these "cognitive linguistics" approaches will also be examined in the next section where we look at theories of metaphor. An overview of some other work on narrative which focuses on cognition (and incidentally on language) is Herman (2003).

9 Genre

Fabb (2002) distinguishes between terms in literary linguistics which are part of a cognitively specialized computation, and those which are best understood as concepts by which we classify literary texts (and their component parts). For example, "sentence" is a term which is part of the syntax, while "Ciceronian sentence" is a term which is a concept by which we classify sentences into kinds but which has no specific cognitive status. Fabb argues that generic terms such as "sonnet" or "tragedy" might best be understood as we understand any concept (i.e., "being a sonnet" is analogous to "being a chair" and presents the same problems of understanding how exactly the concept/category works). That is, the identification of a text as a sonnet or as a tragedy is, like the categorization of anything, a matter of prototypes, or exemplars, or in terms of defining component features (depending on one's theory of concepts; see Murphy 2002). Fabb (2002) argues that genre is assigned to a text by inference based on evidence (sometimes partial or contradictory) which the text provides: It is a communicated form rather than a constitutive form. That is, the text warrants implicatures whose content is a description of the text's genre; the genre holds of the text only as a result of implicature, and is not in any sense an objective fact or inherent quality of the text. This means that its genre is an implied meaning of the text, and like any implied meaning can be ambiguous, indeterminate or vague.

There may be a link between aesthetic experience and (cognitive) epistemology: Seeking and discovery may be a source of aesthetic pleasure, relevant perhaps in the resolution of complications in a narrative, or to other kinds of closure. It is also possible that we gain some aesthetic pleasure from difficulties in discovery or in establishing meaning: Thus, ambiguity, uncertainty, and vagueness have been assigned aesthetic value by literary critics (e.g., Empson 1953) and by contemporary (e.g., poststructuralist) literary theory. Fabb (2002) argues that many kinds of literary form can be understood as concepts/categories, including for example the genre of a text, and that literary texts are often formulated in a way which both invites categorization (e.g., determining genre) and makes it difficult to be certain about categorization.

10 Complexity and Difficulty

One of the characteristics of many literary texts is that their language is subject to additional regulation beyond the regulation imposed by phonology, syntax, etc. Thus the text might be divided into lines which are regulated in a specific way, or there might be a requirement for rhyme, or specific types of lexical or syntactic repetition. Such texts are often culturally valued,

so we might ask whether there is a relation between being valued and having this additional regulation. Richard Bauman (1975) suggests that verbal art is a kind of verbal behaviour whose fundamental characteristic is that it is evaluated by an audience. The text and its performance are valued according to criteria, and rules for language offer such criteria, particularly if they are difficult to achieve. Production of a regulated object by a skilled artist carries some inherent, and apparently universal, cultural value.

These are examples of regulation as a way of creating some difficulty which an author must overcome. But regulation may also have a value for the audience, beyond simply an admiration of an author's skill. Perhaps the cognitive processes which are required in a reader or hearer by the regulated aspects of literature are themselves capable of generating something valued, such as aesthetic experience. Various formal devices such as rhyme and meter restrict the selection space from which lexical items are chosen to decode the perceived sound stream, and it seems that this easing of processing effort has the effect of producing pleasure and a sense of truth and familiarity (Reber, Schwarz, and Winkielman 2004). Thus, when a line must rhyme, the range of lexical items which can be chosen to fill the rhyming slot is restricted: It is easier for the hearer to identify the word, thus reducing processing effort and producing these effects. These are kinds of ease or simplicity, but complexity must also play a role in generating aesthetic experience. Various kinds of complexity can be discovered in the ways that language is regulated in literature; at present there is little evidence that these are correlated with some experiential effect, such that these complexities would have some function which justifies the additional regulation they involve. For example, the term "tension" is used in literary criticism and some literary linguistics to describe the degree to which a text varies from a prototype: An iambic pentameter line has a prototype form of 10 syllables with odd unstressed and even stressed, and an actual line might demonstrate "tension" which can be measured by calculating how far its stress pattern varies from this. Though this is primarily a description of form the term "tension" suggests that there are psychological effects associated with this phenomenon. Issues around various notions of tension are discussed by Miller (1977).

This final section has addressed what I see as the next major challenge for literary linguistics, which is to ask how the regulation of language in literature relates to the value which literature carries, and in particular how it relates either to the adding of value through the compositional process, or to the experience of value in the reception of the text.

SUGGESTIONS FOR FURTHER READING

Banti, G. and Giannattasio, F. (2004). Poetry. In A. Duranti (ed.), *A Companion to Linguistic Anthropology* (pp. 290–320). Oxford: Blackwell.

Greene, R. (ed.) (2012). *The Princeton Encyclopedia of Poetry and Poetics*, 4th edn. Princeton, NJ: Princeton University Press.

RELEVANT JOURNALS AND SOCIETIES

Journals
International Journal of Literary Linguistics
Journal of Literary Semantics
Language and Literature
Oral Tradition
Studia Metrica et Poetica
Style

Societies
Poetics and Linguistics Association

EMERGING TRENDS AND RESEARCH QUESTIONS

1. Psychology of literary language
2. Typology
3. Language and music
4. Statistical poetics
5. Parallelism

REFERENCES

Austin, T. R. (1984). *Language Crafted. A Linguistic Theory of Poetic Syntax.* Bloomington, IN: Indiana University Press.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29.

Bamberg, M. G. W. (1997). Oral versions of personal experience: three decades of narrative analysis. *Journal of Narrative and Life History*, 7 (special issue). London: Erlbaum.

Banfield, A. (1982). *Unspeakable Sentences: Narration and Representation in the Language of Fiction.* Boston, MA: Routledge & Kegan Paul.

Bauman, R. (1975). Verbal art as performance. *American Anthropologist*, 77, 290–311.

Bright, W. (1990). "With one lip, with two lips": parallelism in Nahuatl. *Language*, 66, 3, 437–42.

Dell, F. and Halle, J. (2009). Comparing musical textsetting in French and in English songs. In J.-L. Aroui and A. Arleo (eds), *Towards a Typology of Poetic Forms* (pp. 63–78). Amsterdam: John Benjamins.

Dell, F. and Elmedlaoui, M. (2008). *Poetic Meter and Musical Form in Tashlhiyt Berber Songs*. Cologne: Rüdiger Köppe.

Empson, W. (1953). *Seven Types of Ambiguity*, 3rd edn. London: Chatto & Windus.

Fabb, N. (1997). *Linguistics and Literature. Language in the Verbal Arts of the World.* Oxford: Blackwell.

Fabb, N. (1999). Verse constituency and the locality of alliteration. *Lingua*, 108, 223–45.

Fabb, N. (2002). *Language and Literary Structure: The Linguistic Analysis of Form in Verse and Narrative.* Cambridge: Cambridge University Press.

Fabb, N. (2010). Is literary language a development of ordinary language? *Lingua*, 120, 5, 1219–32.

Fabb, N. (2014). The verse-line as a whole unit in working memory, ease of processing, and the aesthetic effects of form. *Royal Institute of Philosophy Supplement*, 75, 1–22.

Fabb, N. (2015). *What Is Poetry? Language and Memory in the Poems of the World*. Cambridge: Cambridge University Press.

Fabb, N. and Halle, M. (2008). *Meter in Poetry: A New Theory*. Cambridge: Cambridge University Press.

Fitzgerald, C. M. (2007). An optimality treatment of syntactic inversions in English verse. *Language Sciences*, 29, 203–17.

Fludernik, M. (1996). *Towards a "Natural" Narratology*. London: Routledge.

Fludernik, M. (2003). Natural narratology and cognitive parameters. In Herman 2003, pp. 243–67.

Forth, G. (1988). Fashioned speech, full communication: aspects of eastern Sumbanese ritual language. In J. J. Fox (ed.), *To Speak in Pairs: Essays on the Ritual Languages of Eastern Indonesia* (pp. 129–60). Cambridge: Cambridge University Press.

Fowler, R. (ed.) (1966). *Essays on Style and Language: Linguistic and Critical Approaches to Literary Style.* London: Routledge & Kegan Paul.

Fox, J. J. (ed.) (1988). *To Speak in Pairs: Essays on the Ritual Languages of Eastern Indonesia.* Cambridge: Cambridge University press.

Golston, C. and Riad, T. (2000). The phonology of Classical Greek meter. *Linguistics*, 38, 1, 99–167.

Golston, C. and Riad, T. (2003). Scansion and alliteration in *Beowulf*. *Jahrbuch für Internationale Germanistik*, 35, 1, 77–105.

Hanson, K. (1992). *Resolution in Modern Meters.* PhD dissertation, Stanford University.

Hanson, K. and Kiparsky, P. (1996). A parametric theory of poetic meter. *Language*, 72, 2, 287–335.

Hayes, B. (2003 [1989]). The prosodic hierarchy in meter. In P. Kiparsky and G. Youmans (eds), *Rhythm and Meter* (pp. 201–60). Orlando, FL: Academic Press.

Hayes, B. (1995). *Metrical Stress Theory*. Chicago, IL: University of Chicago Press.

Hayes, B. and Moore-Cantwell, C. (2011). Gerard Manley Hopkins's sprung rhythm: corpus study and stochastic grammar. *Phonology*, 28, 235–82.

Hayes, B. P. and MacEachern, M. (1996). Are there lines in folk poetry? *UCLA Working Papers in Phonology*, 1, 125–42.

Hayes, B. and MacEachern, M. (1998). Quatrain form in English folk verse. *Language*, 74, 473–507.

Hendricks, W. O. (1969). Three models for the description of poetry. *Journal of Linguistics*, 5, 1–22.

Herman, D. (ed.) (2003). *Narrative Theory and the Cognitive Sciences*. Stanford, CA: CSLI Publications.

Holtman, A. (1996). *A Generative Theory of Rhyme: An Optimality Approach*. PhD thesis, Utrecht University.

Hopper, P. J. and Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, 56, 251–99.

Hymes, D. (1981). *"In vain I tried to tell you." Essays in Native American Ethnopoetics*. Philadelphia, PA: University of Pennsylvania Press.

Jakobson, R. (1960). Linguistics and poetics. In T. Sebeok (ed.), *Style in Language* (pp. 350–77). Cambridge, MA: MIT Press. (Reprinted in K. Pomorska and S. Rudy (eds) (1990), *Roman Jakobson: Language in Literature* (pp. 62–94). Cambridge, MA: Harvard University Press.)

Kawahara, S. (2007). Half rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics*, 16, 2, 113–44.

Kiparsky, P. (1973). The role of linguistics in a theory of poetry. In *Language as a human problem*. Special issue of *Daedalus*. 102, 3, 231–44. (Reprinted in D. C. Freeman (ed.) (1981). *Essays in Modern Stylistics* (pp. 9–23). London: Methuen.)

Kiparsky, P. (1977). The rhythmic structure of English verse. *Linguistic Inquiry*, 8, 2, 189–247.

Labov, W. and Waletzky, J. (1967). Narrative analysis: oral versions of personal experience. In J. Helm (ed.), *Essays on the Verbal and Visual Arts. Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society* (pp. 12–44). Seattle, WA: University of Washington Press. (Reprinted in Bamberg 1997, pp. 3–38.)

Léon-Portilla, M. (1985). Nahuatl literature. In M. S. Edmonson (ed.), *Supplement to the Handbook of Middle American Indians: Volume 3. Literatures*. Austin, TX: University of Texas Press. 7–43.

Levin, S. R. (1964). Poetry and grammaticalness. In H. G. Lunt (ed.), *Proceedings of the Ninth International Congress of Linguists*. The Hague: Mouton. (Reprinted in S. Chatman and S. R. Levin (eds) 1967, *Essays on the Language of Literature*. Boston, MA: Houghton Mifflin.)

Lotz, J. (1960). Metrical typology. In T. Sebeok (ed.), *Style in Language* (pp. 135–48). Cambridge, MA: MIT Press.

MacMahon, B. (2007). The effects of sound patterning in poetry: a cognitive pragmatic approach. *Journal of Literary Semantics*, 36, 103–20.

Malone, J. L. (1988). Underspecification theory and Turkish rhyme. *Phonology*, 5, 293–7.

Miller, D. G. (1977). Language change and poetic options. *Language*, 53, 21–38.

Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

O'Neil, W. (2001). Grammar games: unscrambling Skaldic syntax. In M. Kenstowicz (ed.), *Ken Hale: A Life in Language* (pp. 339–54). Cambridge, MA: MIT Press.

Pilkington, A. (2000). *Poetic Effects: A Relevance Theory Perspective*. Amsterdam: John Benjamins.

Reber, R, Schwarz, N., and Winkielman, P. (2004). Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8, 364–82.

Sapir, E. (1967). *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and World. (First published 1921.)

Sperber, D. (ed.) (2000). *Metarepresentations: A Multidisciplinary Perspective*. New York: Oxford University Press.

Sperber, D. and Wilson, D. (1985/6). Loose talk. *Proceedings of the Aristotelian Society*, 86, 153–71.

Sperber, D. and Wilson, D. (1995). *Relevance: Communication and Cognition*, 2nd edn. Oxford: Blackwell.

Steriade, D. (2003). Knowledge of similarity and narrow lexical override. In P. Novak and C. Yoquelet (eds), *Proceedings of the Berkeley Linguistics Society*, 29, 583–98.

Tarlinskaja, M. (1976). *English Verse: Theory and History*. The Hague: Mouton.

Thorne, J. P. (1965). Stylistics and generative grammars. *Journal of Linguistics*, 1, 49–59.

Tristram, H. L. C. (1995). Near-sameness in early insular metrics: oral ancestry and aesthetic potential. *Poetics Today*, 16, 3, 445–70.

Tsur, R. (1998). *Poetic Rhythm: Structure Aand Performance: An Empirical Study in Cognitive Poetics*. Berne: Peter Lang.

Turpin, M. (2007). The poetics of central Australian song. *Australian Aboriginal Studies*, 2, 100–15.

Voorhoeve, C. L. (1977). Ta-poman: metaphorical use of words and poetic vocabulary in Asmat songs. *Pacific Linguistics*, C, 19–38.

Yip, M. (1999). Reduplication as alliteration and rhyme. *GLOT International*, 4, 1–7.

Youmans, G. (1983). Generative tests for generative meter. *Language*, 59, 1, 67–92.

Zunshine, L. (2006). *Why We Read Fiction: Theory of Mind and the Novel*. Columbus, OH: Ohio State University Press.

Zwicky, A. M. (1976). Well, this rock and roll has got to stop, junior's head is hard as a rock. Paper presented at the 12th Annual Meeting of the Chicago Linguistic Society.

23 Linguistic Anthropology and Ethnolinguistics

STEVEN P. BLACK AND ELIZABETH A. FALCONI*

1 Introduction

Linguistic anthropology and ethnolinguistics are two labels that have been used to designate ethnographic research about language. Broadly conceived of as scholarship on how linguistic diversity relates to cultural diversity, these disciplinary names index different national academic traditions (ethnolinguistics is a more common designation in Europe and elsewhere, while linguistic anthropology is more common in the US) and a historical split in the United States from Chomskyan approaches to language (designating an affiliation with anthropology and a distancing from linguistics). Due to this historical fission, some American linguists may wonder what one might gain from an ethnographic approach. In the same fashion, some linguistic anthropologists might wonder how anyone could study language without a thorough understanding of cultural context. With an increasing number of linguists expressing an interest in subjects such as language contact, multilingualism, language endangerment, and language obsolescence (topics well suited to ethnographic exploration), the time is ripe to retrace the lineage that these two disciplines share and to explore contemporary connections.

Linguistic anthropology and ethnolinguistics offer theoretical and practical tools for understanding the sociocultural worlds that are inter-related with linguistic practices. The study of language and culture is considered an entire subfield of linguistics and/or anthropology. As a result, it could never be hoped to cover the full depth and breadth of the subdiscipline in a single chapter (for book-length treatments, see Ahearn 2011; Duranti 1997; Enfield, Kockelman, and Sidnell 2014). Some theoretical foci considered to be classic – linguistic relativity and ethnoscience, for example – are left out of the chapter altogether due to a lack of robust recent interest in the topics. Furthermore, while it is common to include linguistic excerpts as examples of particular phenomena, in linguistic anthropology discussion of linguistic practices is usually placed within the quite complex context of specific sociocultural and historical patterns. As a result, it is difficult to include linguistic excerpts in an overview piece such as this, but they are included when they are pertinent and easily accessible. Finally, while the concepts of race, ethnicity, gender, and class are central to most work in the subdiscipline, it has been decided to interpolate these foci within discussions of theoretical frameworks. This chapter overviews core theories utilized in

* This chapter is a result of the equal contribution of both authors.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

contemporary linguistic anthropology and ethnolinguistics, discussing their historical development and suggesting areas for productive dialogue with other work in linguistics.

1.1 Naming as disciplinary (and social) practice

What is in a name? A great deal, if you ask a scholar of language and culture (e.g., Basso 1988; Rymes 1996). Among linguistic anthropologists in the United States, one would be hard pressed to find a junior scholar that knows much of anything about the term ethnolinguistics. Though sometimes used in the first half of the twentieth century, in the late 1960s the term was abandoned for theoretical reasons (for in depth historical accounts see Duranti 2003; Hymes 1963; Rumsey 2013). In 1964, in the journal *American Anthropologist*, Karl Teeter wrote a brief communication on the terms "anthropological linguistics" and "linguistic anthropology," arguing that the latter term was more appropriate for "the study of language data and the forms of grammars and lexica with a view to the elucidation of the relations of language and culture (or the place of language *in* culture)" (Teeter 1964: 878–9). Not once did Teeter use the term ethnolinguistics. In a response to Teeter published the next year, C. F. Voegelin (1965) chided Teeter for not mentioning the sociology of language, or sociolinguistics. Voegelin contrasted sociolinguistics with ethnolinguistics "as practiced by anthropologists" (p. 484). It seems that he may have been using ethnolinguistics as a derogatory term designating anthropological research on language by scholars without proper linguistics training (Duranti 2003: 325, fn.6).

Voegelin (1965) was also looking toward the coming intellectual tidal wave of Noam Chomsky's (1964) transformational-generative grammar. Chomsky's theory pushed the expansion of linguistics departments in the United States (see Chapter 6). It also precipitated a split between scholars interested in the underlying universal human language faculty posited to generate particular languages, on the one hand, and scholars who viewed language as necessarily embedded in (and thus to be studied within) social and cultural context (variously defined), on the other hand (Rumsey 2013: 273). A number of key scholars rejected the term anthropological linguistics in favor of linguistic anthropology to differentiate their work from the cognitive/psychological turn in linguistics and to emphasize connections with the ethnographic study of culture (Duranti 2003). While this designation has helped scholars of language and culture to promote a distinct disciplinary vision, it is also important to note that, throughout the latter twentieth and early twenty-first century, many researchers have continued to engage in productive dialogue with various subdisciplines in linguistics and beyond (Silverstein 2006).

The term ethnolinguistics continues to hold sway in other national scholarly traditions (Rumsey 2013: 275). In some of those national traditions it is also used to refer to the use of language to construct ethnic identities (e.g., Webb 2002: 152). In this chapter, the name linguistic anthropology is used but the term ethnolinguistics is also retained to highlight three elements of interconnection: (1) a historical kinship with subfields in linguistics; (2) an ongoing exchange with scholarly traditions in European countries, Australia, and beyond, especially in regions of the world not traditionally known as scholarly hubs; and (3), increased efforts to reconnect linguistic anthropologists with their colleagues stationed in linguistics departments.

2 An Ethnographic Approach to Language

Many linguistic anthropologists trace their approach back to the ethnographic tradition developed by Franz Boas and his students as well as to the "descriptive linguistics" of Leonard Bloomfield, among others (Rumsey 2013: 284; Stocking 1966). It has been indicated that these traditions may have been antigeneralization and antitheorization at one time (see Chapter 6). However, it seems that it is simply that the character of generalizations and focus of theorization is distinct from that of other types of linguistic study. Most linguistic anthropology research projects are ethnographic

in nature. This includes extensive participant observation, which is usually complemented by audio-video recording and transcription / translation of recordings. Longitudinal research is heavily favored. Researchers typically spend a year or more at their chosen field sites in order to more fully grasp the complex ways that everyday linguistic practices are bound up with other domains of social life. This methodology distinguishes linguistic anthropology and ethnolinguistics from field linguistics (Chapter 4), which involves elicitation, interviews with native speakers, and the recording of oral texts collected during short-term fieldwork trips. At its best, participant observation results in qualitative holistic understandings of how a select number of people go about their daily lives. Through these methods, ethnography provides insights about the intersection of language with class, gender, and race, among other social, historical, economic, and political patterns.

Participant observation is deceptively simple: A researcher observes social life in a given context, taking copious fieldnotes, and she also participates to the best of her abilities in order to better grasp the complexities of cultural practice and lived experience. A great deal of scholarship explores the significance of ethnography within the social sciences, the practical issues involved in ethnographic fieldwork, and the epistemological limits of this form of knowledge production (e.g., Bernard 2011; Jackson 1998; Clifford and Marcus 1986). While the US tradition of ethnography through participant observation is, as mentioned, essentially Boasian, many would point to British social anthropologist Bronislaw Malinowski as a forebear of modern ethnographic methods. Malinowski arrived at the Trobriand Islands to conduct anthropological research just prior to the outbreak of World War I. As a result of the war, he remained trapped there for several years. This accident of history cemented the research method of living with a given community for a period of a year or more with the aim of learning to learn the totality of a particular "culture." About language and culture, Malinowski wrote, "Language is intertwined with the education of the young, with social intercourse, with the administration of law, the carrying out of ritual, and with all other forms of practical co-operation" (Malinowski 1935: vii). This approach to language emphasized a form of functionalism – focusing on the function of language in social life.

In more recent work, the notion of homogeneous, timeless, and discretely bounded "cultures" employed by Malinowski and his contemporaries is now understood as heterogeneous, historically situated, and largely unbounded (e.g., Abu-Lughod 1991; Latour 2005). The researcher is no longer viewed as someone with a preternatural ability to empathize with research participants (Geertz 1983). However, Malinowski's functionalist view of language (broadly understood) is still utilized. Furthermore, ethnography rooted in participant observation remains a significant method for the production of holistic qualitative knowledge about everyday life and cultural practices, and language remains the central medium through which ethnographic research is carried out (Gal 2012; Marcus 2008).

2.1 A note on recording and transcription

As noted, a key part of many contemporary ethnographic studies of language is the intensive use of audio-video recording and subsequent transcription of "naturally occurring" speech. This type of data collection contrasts with (but does not necessarily conflict with) sociolinguistic interviews and linguistic elicitation. Here, the word "naturally" should not be seen as a claim that speakers are unaffected by the presence of researcher or recording device. Instead, the term simply indicates the audio or audio-video recording of communication as people as go about their social activities. Disambiguating the impact of a recording device on people's actions is just a special case of what is known as the *participant observer's paradox*. The paradox is that to observe cultural practices a researcher must be present, but by being present the researcher necessarily alters the performance of those cultural practices. This paradox is not an insurmountable epistemological barrier to research. Rather, it is a reminder that researchers should always account for the various ways in which their data collection methods and their very presence impact the data that is collected (Duranti 1997: 118). Adding a recording device means that researchers must

attend to both how their own presence and how the use of a recording device changes the content and form of ongoing social activities.

After conducting participant observation supplemented by the recording of a corpus of data, the linguistic anthropologist or ethnolinguist sits down with a research assistant (often a person or people from the community of research participants) and begins the arduous task of creating preliminary transcripts and translations of (sometimes hundreds of) hours of recorded data. Often these transcripts are annotated with research assistants' commentary on cultural and linguistic practices (Schieffelin 1990). Here, it is emphasized that ethnography, like culture itself, emerges out of dialogue (Tedlock and Mannheim 1995). Like all documentation of actual speech, (annotated) transcription results in only a partial representation of the wealth of semiotic resources (e.g., facial expression, prosody, gesture, tone) brought to bear in an interaction. The researcher's choice to highlight some communicative features at the expense of others is thus a theoretical decision with practical and political implications (Bucholtz 2000; Ochs 1979). At the same time, this partial nature of transcripts makes them useful for identifying and isolating linguistic patterns amid the deluge of semiosis that is even a few moments of face-to-face communication. In creating and revising transcripts, one must always remember that a transcript is not a full representation of a recording, and indeed, a recording is not a complete documentation of the human experience of social life (Duranti 2006). Recognizing these epistemological uncertainties, many scholars analyze the place of language in social life by triangulating transcripts of recordings with participant observation, copious fieldnotes, and, when available, the previous research of sociocultural anthropologists and others (Philips 2013).

2.2 *Communicative competence, speech communities, and indexicality*

Ethnographic approaches to the study of language have resulted in a number of key research topics and questions that will be explored below, including language socialization, language ideologies, language endangerment and obsolescence, and verbal art. These topics are approached from a theoretical standpoint that is distinct from other perspectives on the study of language – that language should be evaluated in terms of whether or not it is *successful* in producing social action rather than simply in terms of whether or not it is grammatically *correct*. Here, three key ideas that underlie this theoretical standpoint are briefly outlined: communicative competence, speech communities, and indexicality.[1]

The concept of communicative competence was developed just after, and in explicit opposition to, some aspects of the Chomskyan paradigm shift that took hold in American linguistics in the late 1960s and early 1970s. Specifically, communicative competence was precipitated by Chomsky's redefinition of the terms competence and performance. Chomsky (1964) famously understood performance – or actual talk – as entwined with "grammatically irrelevant conditions [such] as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying [the ideal speaker-listener's] knowledge of the language" (pp. 3–4). This suggests that research using speech data uttered in the context of social activities is hampered, if not derailed entirely, by the detritus of human imperfection. For generative grammar, then, the proper focus of study must be competence, the underlying cognitive or mental knowledge that generates grammatically correct language.

John Gumperz and Dell Hymes (1964), proponents of the *Ethnography of Communication* research paradigm, countered that Chomsky's vision of linguistic competence is only one part of what people must know about a language to be a competent speaker. Hymes (1972) argued that people must be able to master four aspects of what he termed *communicative* competence:

1 Whether (and to what degree) something is formally possible
2 Whether (and to what degree) something is feasible by virtue of the means of implementation available

3 Whether (and to what degree) something is appropriate (adequate; happy, successful) in relation to a context in which it is used and evaluated
4 Whether (and to what degree) something is in fact done, actually performed, and what its doing entails. (Hymes 1972; see also Goodwin 2004)

The notions of feasibility, appropriateness, and entailment (parts 2, 3, and 4 of Hymes' model) have remained mainstays of linguistic anthropological research since this groundbreaking piece was published. Scholarship has explored how grammatical structuring is inter-related with hierarchy, gender, and formality, among other aspects of social life (e.g., Irvine 1979; Philips 1991; Ochs 1992; Silverstein 1985). Part 1 of the model, on the other hand, remains underexamined. Though there are many examples of excellent ethnographic work that includes descriptive grammars of undocumented languages, in general scholars of language and culture tend to rely on previously published descriptive linguistic work to answer "whether (and to what degree) something is formally possible."

A key orienting principle for the Gumperz and Hymes' ethnography of communication paradigm is the notion of the speech community (see also Chapter 21). In contrast with Chomsky's assumption of a homogeneous community of ideal speaker-listeners, Gumperz and others studied linguistic variation within what was seen as a community of speakers (e.g., Gumperz 1964). This study of linguistic difference and its social meaning within and between communities became central to studies in variationist sociolinguistics (e.g., Labov 1966; see Chapter 25). It has been productive to think about speech communities in this way, with speakers divided into distinct (if heterogeneous) groups, or "communities." Still, in many cases it is empirically difficult to determine where one community ends and the next begins. More recent work complicates this classic model, drawing from notions such as "communities of practice" and/or speech network theory (Bucholtz 1999; Dorian 1992; Morgan 2004). In contemporary research the concept of community has proved most useful for understanding situations in which people imagine or believe themselves to be part of a single community despite considerable linguistic variation (e.g., Falconi 2013; Spitulnik 1996).

In addition to communicative competence and speech community, theorization of indexicality is a central part of scholarship on language and culture. Michael Silverstein produced formative work on this topic, synthesizing the perspectives of linguist Roman Jakobson (e.g., 1960) with that of philosopher and founder of American pragmatism Charles Peirce (1985 [1897]). A key tenet of both Jakobson and Peirce's work is that language is greater than its mere referential or symbolic content. In fact, as Silverstein argues, "the vast residue of language is culture, and culture is pragmatic" (Silverstein 1976: 54). Where classic linguistic theory states that the relationship between sound and meaning is *arbitrary*, a Peircean perspective counters that the relationship is *conventional* (cf. de Saussure 1916). The term "arbitrary" would imply that the choice between interchangeable linguistic forms has no history or consequences; yet much of the communicative force of language comes from indexical associations that are reproduced or transformed in moments of talk (and thus become conventional or symbolic).

Understanding language as conventional (rather than arbitrary) and saturated by indexicality is essential for understanding the *why* of language: why diachronic shifts in language occur, and why people make particular choices in (synchronic) language use. For instance, linguists have long emphasized that the language variety often called African American English (AAE) is rule governed and grammatical. Despite this fact, in mainstream contexts the use of features of AAE has a pragmatic force that is usually but not always evaluated negatively by white middle-class Americans. In contrast, speakers of AAE may evaluate the same linguistic features significantly more positively in communication with other speakers of the language variety (e.g., Alim and Smitherman 2012; Labov 1972). Theorization of indexicality has allowed researchers to explain such patterns of language and culture, exploring the specific ways that linguistic structures are interrelated with both the pragmatic force of language and people's feelings about structure and pragmatics (e.g., Hill 1998; Ochs 1992; Silverstein 1981).

2.3 Embodied participation in conversation

In addition to these theoretical perspectives on communicative competence, speech community, and indexicality, contemporary ethnographic research on language recognizes the central role that multiple semiotic resources or modalities, including the artifacts and the material world, play in communication (Goodwin and Goodwin 2000; Keane 2007). In much of linguistics, semiotic modalities such as gesture, facial expression, body orientation, eye gaze, human artifacts, the built environment, and even conversational organization are pushed aside in an effort to isolate "core" elements such as phonology, morphology, and syntax. In actual speech, however, embodied semiotic resources are crucial for constituting the meaning of words and coordinating social activities. In this area of research, especially, linguistic anthropology overlaps considerably with conversation analysis (a subfield of sociology closely related to discourse analysis – see Chapter 21). Conversation analysis developed from ethnomethodology – the study of everyday interactions and their sociological significance (Garfinkel 1967; Heritage 1984; Sacks 1984). Conversation analytic approaches investigate conversations as they temporally "unfold" (see Sidnell and Stivers 2012). This yields a window into what might be termed the grammar of interaction (Sacks, Schegloff, and Jefferson 1974). Research on embodiment drawing from conversation analysis discusses how interlocutors co-construct meaning in real time while utilizing multiple semiotic resources (e.g., eye gaze, facial expression, gesture). All grammar, not just conversational grammar, supports this function of talk (Raymond 2009).

From such a perspective, interactions cannot be understood using the traditional speaker-listener model in which speakers take turns actively talking while other parties passively listen. Rather, all parties are active co-participants in temporally unfolding social activities. Even when not speaking, participants employ a myriad of embodied resources that shape the flow of conversation. Like other semiotic resources, gesture is an intersubjective (communicative) tool, not merely a "tell" indicating hidden psychological preferences (Haviland 2004). For instance, as a person speaks, coparticipants' shifts in eye gaze can lead the speaker to change the content of even a single sentence midstream (Goodwin 1979; 1981). From this viewpoint, the word "language" is an inadequate moniker for the complex symphony of multiple semiotic resources that multiple participants deploy in even a few moments of face-to-face interaction (see Streeck, Goodwin, and LeBaron 2012). Furthermore, in contrast to many linguists' dismissal of pauses, false starts, restarts, and overlaps, this stream of research has revealed that these "mistakes" may perform vital conversational functions (e.g., Goodwin 1979; Schegloff 2000).

Modeling how embodied language is linked to sociocultural distinctions, the concepts of communicative competence, speech communities, and indexicality underlie most, if not all, scholarly investigations into language and culture. In conjunction with ethnographic fieldwork methods, these and other key ideas provide a powerful toolkit for studying the following topics to be discussed below: how language acquisition is inexorably connected to socialization into cultural patterns (language socialization); how language is integral in the reproduction of power and inequality (language ideologies); how languages shift over time, and how people resist language obsolescence (language endangerment); and how language is framed in particular ways to exploit communicative ambiguity and produce ritual efficacy (verbal art).

3 Language Socialization

In the 1980s and 1990s, scholars drew on multidisciplinary training in anthropology, developmental psychology and psycholinguistics, descriptive linguistics, and ethnomethodology to develop theoretical perspectives on language socialization. In response to previous anthropological research on communication across cultures that had focused disproportionate attention on the linguistic and cultural activities of adult men, initial language socialization research

focused on the language practices of caregivers (predominantly women) and children. The language socialization approach emphasizes longitudinal research amongst novices and experts in recurring familiar contexts in order to track the acquisition of cultural and communicative competence alongside processes of social and cognitive development. Such research emphasizes that enculturation and language acquisition are actually two parts of the same process – socialization to language and socialization *through* language (into communicative competence).

Bambi Schieffelin and Elinor Ochs (1986) developed the theory of language socialization based on ethnographic research on the young children's acquisition of language and cultural competence in small tight-knit communities in Papua New Guinea and Samoa. Following a wave of research on the myriad ways that adults systematically alter their language use patterns to accommodate children and facilitate children's understanding and use of language (e.g., Phillips 1973; Sachs, Brown, and Salerno 1976), Schieffelin and Ochs focused special attention on the cultural patterning of such accommodations and practices. They asserted that "language socialization begins at the moment of social contact in the life of a human being" (1986: 164). This theoretical framework is grounded in the following two tenets:

1 The process of acquiring language is deeply affected by the process of becoming a competent member of society.
2 The process of becoming a competent member of society is realized to a large extent through language, by acquiring knowledge of its functions, social distribution, and interpretations in and across socially defined situations, i.e. through exchanges of language in particular situations. (Ochs and Schieffelin 1984: 277)

Another way to think about this is that language socialization includes socialization *to* language (previously described as language acquisition) and socialization *through* language into particular cultural practices (previously described as enculturation). One way to frame this would be to say that language socialization research investigates the acquisition of communicative competence, providing support for Hymes' insight that linguistic competence is inextricably bound up with the social and cultural contexts within which speakers are located.

For instance, one aspect of language socialization that Ochs (1982) studied was the acquisition of linguistic markers of ergativity within an age and status-based hierarchical social context. Samoan is an ergative-absolutive language in which nominative case marking linguistically encodes culturally specific notions of agency (Duranti 1990). At the level of Samoan linguistic structure, the difference between agents – entities that act on other entities – and nonagents is fairly straightforward. However, Ochs demonstrated that (1) Samoan adults variably deployed the ergative marker depending on the perceived or desired social distance between speaker and addressee; and (2) as a result, Samoan children who were not present in formal social situations far from close kin were unlikely to hear and thus unlikely to acquire the ergative marker at a young age. Children aged 2–4 did recognize the ergative-absolutive distinction but tended to indicate it through word order rather than marking it with an ergative affix. Here is an example in which a child, age three, alternates between word order and morphological ergative marking as he tried to scare his peers about the threat of a tractor coming their way:

N: *sua* *okoul*
Strike down you (PL)
'It's going to strike you down.'

sua *okou e* *makoa* *lolil*
strike down you ERG our(EXCL) truck
'Our truck is going to strike you down.'

sua *okou makou* *loli* *foul*
strike down you our (excl) truck new
'Our new truck is going to strike you down.' (adapted from Ochs 1982: 656)

This research demonstrates that language acquisition is meaningfully impacted by the social salience and social distribution of morphological features. It also suggests that classification of languages might meaningfully include a sociological dimension, recognizing that in some languages the use of particular linguistic features (such as ergative markers) is shaped by distinctions in social relationships (Ochs 1982: 665–7).

3.1 From linguistic input to interactional scaffolding

From a language socialization perspective, every social interaction offers the opportunity to create and negotiate shared understandings, perspectives, and worldviews. Interactions involving novices and more competent members of a social group also provide a window into how culturally specific ideas about knowledge, prestige, and social power impact the socialization process. It bears mentioning that this approach differs from the study of first and second language acquisition, which focuses on the investigation of linguistic competence, how language comprehension is organized and produced, as well as how it is defined at different stages of human development and the life cycles of individuals (see Chapters 19 and 27). Ochs and Schieffelin (1984) offer an anthropological response to prevalent debates, fanned by Chomsky's scholarship, regarding the respective roles of nature (innate biological mechanisms) versus nurture (environmental and cultural influences) in the acquisition of linguistic competence. Prominent behaviorist scholars asserted that children came into the world effectively *tabula rasa* and acquired language through stimuli provided by adults. Chomsky (1959), on the other hand, argued that children were possessed of an innate human cognitive capacity (the language faculty) that predisposed them to "generate" grammar from the stimuli to which they were exposed in daily lives (see also Pinker 1994). While not discounting the notion of generative grammar itself, empirical long-term ethnographic research on language socialization revealed that the Chomskyan paradigm too tightly circumscribes the role of adult input to the provision of "relevant information that allows the child to select from Universal Grammar those grammatical principles specific to the particular language that the child will acquire" (Schieffelin and Ochs 1984: 279).

Broadening the scope of inquiry, language socialization scholarship analyzes how experts / adults provide children / novices with interactional scaffolding (Vygotsky 1978) that guides them toward the acquisition of locally valued skills, cultural knowledge, and perspectives. This enables children / novices to move toward collaborative and independent action. Most of the knowledge that speakers typically bring to bear in social activities is tacit, shaping and guiding behavior beneath the level of conscious awareness (see Giddens 1979). However, in the context of socialization this knowledge is often made explicit to facilitate the acquisition of culturally valued practices by children and other novices (see Garrett and Baquedano-Lopez 2002). At the same time, many practices previously presumed to be essential to the language acquisition process, such as baby talk (a register of speech to babies characterized by simplification), are in fact culturally specific. Baby talk and other forms of linguistic accommodation are not universal or necessary for the development of communicative competence.

In some contexts, children are expected to repeat complex adult speech forms with the idea that comprehension will eventually follow. For example, comparative studies of Samoan, Kaluli, and Athabaskan communities have demonstrated that (morphosyntactic) simplification is not a part of adult speech to children in all cultural contexts (Schieffelin and Ochs 1986: 173). Children's own utterances are responded to in a variety of ways, sometimes ignored, sanctioned, or paraphrased in a culturally and linguistically appropriate way. Analysis of these caregiver responses yields a continuum of communicative accommodation from highly child centered (in

which caregivers shape their actions and speech practice to accommodate the limited competency of children) to highly situation centered (in which children are expected to accommodate their actions and utterances to the particular demands of social contexts and events) (Schieffelin and Ochs 1986: 174; see also Meek 2010).

3.2 Socialization throughout the life cycle

Language socialization is a theoretical perspective that can be applied to the study of most communicative acts and social activities. It has been used to explore diverse topics such as morality, exclusion, language shift, the cultural organization of attention, and verbal improvisation (e.g., Fader 2009; Garcia-Sánchez 2014; Garrett 2005; Brown 2012; Duranti and Black 2012). Current research in the field emphasizes two inter-related themes: (1) language socialization is bidirectional; and (2) language socialization continues throughout the life cycle and is thus is not exclusively connected with infants and young children.

First, despite the varying degrees of cultural knowledge and competence that they may possess, children and novices are active participants in every interaction that comprises the socialization process. In some cases, children and novices may even surpass their elders and other culturally recognized "experts" in particular sociocultural domains. For example, the children of migrants often attain greater fluency and/or literacy in the dominant language of the host society than their parents, acting as translators and mediators for them in interactions with individuals from outside of the home community (see Schieffelin and Cochran-Smith 1984; Garrett and Baquedano-Lopez 2002). The active role of children and other novices in language socialization becomes particularly significant when examining situations of language contact and language endangerment as children internalize and react to subtle cues about the relative socioeconomic value of different language varieties.

Second, while initial work on language socialization focused on interactions between caregivers and children, subsequent scholarship has emphasized the ongoing nature of socialization. This is particularly evident in cases where groups of individuals create a shared code in adolescence or adulthood. For example, deaf individuals born into nonsigning families and communities in Nicaragua, highland Chiapas, Mexico, and elsewhere have fueled the emergence of new sign language varieties (see Senghas 1997; Senghas and Monahan 2002; Haviland 2011). More broadly, individuals and social groups across global contexts are continually influenced by contact between diverse linguistic and cultural systems, ideologies and practices, including language standardization, official language policies, and the introduction of literacy into communities without a previous literacy tradition, which will be discussed further below.

4 Language Ideologies

Like language socialization, the theoretical framework of language ideologies expands the range of linguistic inquiry beyond the confines of the analysis of linguistic structure. The concept of language ideologies theorizes the inter-relationship among (1) language structure, (2) language use, and (3) culturally specific systems of value, interpretive frameworks, and worldviews that speakers hold as members of a particular community or social group (Kroskrity 2004). While some of these cultural logics are explicitly stated, many of them are accessible only through extensive ethnographic research on the quotidian behaviors of speakers across contexts (Irvine 2012; Woolard 1998a). Language ideologies are multiple and are embedded in heterogeneous social and cultural practices of various groups. The concept of language ideologies provides a conceptual bridge between fine-grained phonological, morphosyntactic, and/or discourse analysis on the one hand, and the study of the social, historical, political, and economic contexts within which people are speaking and acting, on the other hand (see Irvine 1989).

While the theory of language ideologies has been lauded as a point of connection to social theory (discussed below), less discussed are the connections of the theory to linguistics. Michael Silverstein's first articulation of the term "linguistic ideologies" was, in part, a critique of linguistic structuralism (Silverstein 1979). Building on his work on indexicality (discussed earlier), Silverstein outlined a theory of linguistic ideology that analyzed people's beliefs about the *multiple* functions of language, especially how beliefs about the referential capacities of language intersected with beliefs about the pragmatic force of language (Silverstein 1979; see also Jakobson 1960). His analysis of metapragmatics – explicit or implicit discussions about the forms of social action that speakers accomplish through speaking – was to become a key part of future work on language ideologies (see Silverstein 1993).

A key contribution of language ideologies work is an explanation of language change that spans the social and the linguistic. Silverstein demonstrated that people formed beliefs, attitudes, and emotions about language structure and use based on what was only a partial awareness of structure and use. Using examples from multiple cultural and linguistic contexts, he suggested that certain aspects of linguistic structure are more readily available to folk analysis (and thus available for language ideologies) than others (Silverstein 1981; see also Rumsey 1990). As social factors (e.g., ideologies of language "correctness") motivate people to examine and change their linguistic practices, certain aspects of linguistic practices remain more or less invisible (see Preston 1996).

The concept of language ideologies makes it possible to explain how certain linguistic structures gain prominence and are diffused through multiple communities of speakers. For instance, ideologies research suggests that the incorporation of click consonants into isiZulu is the result of historical borrowing from Khoisan languages for use in a respect / avoidance register (*hlonipha*). Scholars have long been fascinated by click sounds and noted their diffusion across otherwise unrelated languages (e.g., Boas 1911: 45). Historical linguistics research confirms that palato-alveolar, dental, and lateral click consonants were borrowed from Khoisan languages into the *hlonipha* respect / avoidance register (Herbert 1990a, 1990b). This register provided brides and other people in subordinate positions a way to display respect by avoiding particular words, notably including people's first and last names (Finlayson 2002). It has been suggested that clicks provided a simple way to perform this avoidance – by inserting a click consonant a speaker was no longer pronouncing a name to be avoided (Herbert 1990a).

Here, language ideologies scholarship explains what historical linguistics cannot, namely *why* click consonants in particular were chosen (Irvine 1998; Irvine and Gal 2000). At the time of this phonological borrowing, some Khoisan speakers lived on the edges of Bantu-speaking societies, existing in a state of marginalization. The incorporation of click consonants into isiZulu involved three dichotomies: (1) the most prominent and distinct sounds of Khoisan languages (clicks) versus everyday isiZulu; (2) the social position of Khoisan speakers relative to isiZulu speakers; and (3) the social position of speakers of the *hlonipha* respect / avoidance register relative to addressees.

Recursive dichotomies across social and linguistic scales

Phonological:	Khoisan click consonants	vs.	isiZulu speech
$Social_1$:	Khoisan speakers	vs.	isiZulu speakers
$Social_2$:	*Honipha* register speakers	vs.	*Hlonipha* addressees

Khoisan speakers occupied a subordinate position relative to isiZulu speakers, making click consonants an ideologically salient choice for isiZulu speakers themselves to use to index subordination (Irvine and Gal 2000). An ideological evaluation of how phonological difference related to social difference (between Khoisan and isiZulu speakers) was recursively employed to (re)produce another kind of social difference (within communities of isiZulu speakers). Eventually, this led to the incorporation of click consonants into many isiZulu words. This historical language

shift cannot be fully understood without an understanding of the sociocultural dynamics of language use, especially an understanding of language ideologies. This example demonstrates the constant interplay between language structure, language use, and metapragmatics.

4.1 Ideologies, institutions, and social power

A central concern of research on language ideologies is how linguistic practices operate within social institutions (e.g., governments, schools, courtrooms, clinics and hospitals, publishing and media industries). In such settings, power, control, and authority may determine the shape of quotidian behavior and communication. Scholars have demonstrated that bureaucrats' (e.g., court reporters, police officers) control over the structure and content of interactions with laypersons is a form of social power that enables them to shape "the definition of reality in the bureaucratic setting" and "[to impose] a bureaucracy-specific interpretive perspective" (Philips 1998: 212; see also Conley and O'Barr 1990). Here, the language ideologies concept draws from social theory that views social structures and processes of cultural reproduction as products of interactions between individuals and institutions (see Bourdieu 1977, 1991; Gramsci 1971; Foucault 1972, 1980; Williams 1977). As such, the language ideologies concept has allowed for greater engagement with research in sociocultural anthropology and sociology.

Of especial interest has been the role of nationalism and nation-state formation in circulating ideologies that promote an isomorphic relationship between specific languages, cultural traditions, and national belonging. Such ideologies have a naturalizing effect. For example, it is often presumed that to be a German citizen means to speak, read, and write a particular variety of German, rather than French, Polish, or Dutch (to name a few nation-states that share a border with Germany). This naturalizing discourse erases the complexity of the diverse historical and political processes that have resulted in a particular nation-state formation. Such naturalizing discourses suppress particular forms of ethnic, cultural, and linguistic diversity, as well as regional struggles for autonomy and the like (see Irvine 2006, 2012).

The clearest manifestations of the role of language in the nation-building project are language planning policies, meaning deliberate efforts to shape the acquisition, use, and purview of particular languages (Hinton 2012; Jaffe 1999). This includes language standardization and the promotion of official languages (see Chapter 33). Linguistic anthropologists working throughout Europe and in postcolonial contexts have demonstrated how language standardization and the creation of official languages is associated with a political economy of language in which access to and fluency in standard varieties is closely associated with economic and social mobility (see Irvine 1989). In educational contexts, in particular, the value of standard language varieties is both assumed and perpetuated alongside prescriptive language ideologies that correlate standard varieties with grammatical "correctness" (see Cameron 1995; Lippi-Green 1997; Milroy 2001; Preston 1996). Educational institutions in turn depend upon an entire industry of language professionals, grammarians, and publishing companies that produce, print and circulate the dictionaries and grammar textbooks upon which curricula are based, and which constitute secondary and tertiary sites of language ideological production (see Philips 2000).

Descriptive linguists and linguistic anthropologists are active participants in the language industry. We are scholars whose technical training and expertise is often thought to afford a more nuanced awareness of grammatical structure than that of speakers themselves, as discussed earlier. This is particularly true of researchers involved in descriptive linguistics who produce orthographies for previously unwritten languages, a process which constitutes a form of incipient standardization (see Schieffelin and Doucet 1998; Jaffe 1999; French 2003). Such industries and institutions are intimately bound up with the creation of particular forms of metalanguage, or explicit discourse about the form and content of language itself. As Mannheim elucidates, metalanguage:

> carries the process of reproduction of the linguistic code into the consciousness of speakers, where, for better or for worse, it becomes subject to ideological tampering (Silverstein 1979), to the "gentle violence" (nonetheless very real!) of prescriptive traditions and authoritarian control of the categorical systems we live by. (Mannheim 1984: 53)

The "gentle violence" to which Mannheim refers is perhaps clearest in those contexts in which speakers' uptake of a standard language ideology is associated with language shift, or, in an increasing number of cases, language endangerment and obsolescence (discussed below).

4.2 Multilingual contexts and socialization into ideologies

Language ideologies perspectives are powerful tools for the analysis of multilingual contexts. Much recent research within linguistic anthropology is based on ethnographic fieldwork in culturally and ethnically diverse urban settings that often include multilingual social contexts (which are the global norm, see Chapter 26). In such social arenas, linguistic and cultural contact is associated with bilingualism, multilingualism, code-switching, codemixing, bivalency, language shift, and/or obsolescence (Gal 1987; Gumperz 1982; Woolard 1998b). Histories of colonialism have shaped many of these contexts. Oppression has inscribed systemic political and social inequality into relationships among speakers of different languages.

Such settings are fertile ground for research drawing from both language socialization and language ideologies. Both theoretical frameworks facilitate investigation of how speakers in heterogeneous sociolinguistic contexts acquire, understand, and evaluate particular linguistic varieties among the wide array of available linguistic resources. In fact, several scholars helped to develop the theoretical and empirical foundations for study in both language socialization and language ideologies (e.g., Schieffelin, Gal, Woolard) and much subsequent research has integrated both frameworks. In Papua New Guinea, for instance, widespread social changes including Protestant evangelical missionization, the introduction of literacy, and the introduction of capitalist forms of accumulation have impacted language acquisition in the tiny village of Gapun, a community of less than two hundred individuals. These social changes instigated a shift away from the vernacular language Taiap toward the regionally widespread creole language, Tok Pisin. When researcher Don Kulick (1992) asked parents why their children were no longer learning Taiap, they responded, "It's the children. They're all bikhed [strong-willed]." Kulick discovered that a cultural distinction between *hed* (willfulness, selfishness, backwardness) and *save* (enlightened, social, cooperative, worldly), indexed in pre-Christian linguistic practices through distinct registers of Taiap, had been ideologically mapped onto the distinction between Taiap and Tok Pisin. Children, desiring to become *save*, attended to (and eventually mastered) Tok Pisin (Kulick 1992). This demonstrates that children are active participants in the language socialization who attend closely to caregivers' embodied and displayed language ideologies.

Another example from multilingual contexts of how language socialization and language ideologies are inter-related is a pattern language shift in an English-Kaska bilingual community in the Canadian Yukon. There, language shift is bound up with language socialization practices that emphasize "taciturnity and the inappropriateness of very young children as conversational partners" (Meek 2010: 56). At the same time, English dominance, an enduring effect of British colonialism, has become increasingly pervasive in the community. English overwhelms the very limited input in Kaska that children receive from fluent adult speakers. By the time they are old enough to be socially recognized as interlocutors, Kaska youth have not acquired sufficient knowledge of Kaska to communicate fluently with their elders. These examples indicate that language socialization and language ideologies can be understood as complementary theoretical perspectives (Garrett 2004; Garrett and Baquedano-Lopez 2002). The explanatory power of a theoretical synthesis of the two is especially apparent in contexts of language endangerment and obsolescence, as will be discussed below.

5 Language Contact

Social theorists and scholars across disciplines have been captivated by a fundamental tension between processes of cultural reproduction and transformation; or, to put it another way, how it is that humans are able to pass on and retain cultural traditions and practices amid the passage of time and the constancy of social change. Language varieties are inextricably bound up with the reproduction of cultural practices and traditions. As such, they are equally subject to change and transformation over time. This is particularly evident when speakers of distinct languages and/or cultural groups with divergent ideologies of language come into contact and begin to influence one another's behaviors. Largely due to the "multiplicity and contention" (Gal 1998) inherent in all language ideologies, instances of language contact can yield a wide array of possible outcomes. This includes the formation of new languages, bilingualism, multilingualism (see Chapter 26), or language shift. Language varieties not bolstered by state sponsorship or official status – especially nonstandard linguistic varieties and languages spoken by a limited number of people – are particularly prone to rapid processes of shift, and in some cases endangerment or obsolescence.

Instances of language shift and obsolescence are well documented in colonial and postcolonial contexts in which official language policies and language standardization led to the forced imposition of dominant languages (and accompanying dominant language ideologies) in schools and other state-sponsored institutions. The vast majority of languages spoken globally prior to the era of European colonial expansion have died, become dormant, or are categorized as endangered due to the small number of remaining speakers (see Dorian 1992; Hill and Hill 1986; Kroskrity and Field 2009; Kroskrity 2012). In postcolonial contexts worldwide, many indigenous communities continue to experience the cumulative impact of centuries of prejudice targeting speakers of indigenous and/or minority languages. Such prejudice has been built on ideological dichotomies in which indigenous cultural and linguistic practices are categorically opposed to notions of modernity, progress, and economic mobility. Here ethnographic research on language endangerment, shift, maintenance, and revitalization is overviewed.

5.1 Contact, shift, and endangerment

In situations of language contact, idiosyncratic dynamics of particular social, geographic, political, and historical contexts have led to an extraordinary range of outcomes in terms of the relative vitality of the multiple language varieties in speakers' repertoires. Contact-induced change often proceeds at an uneven pace. As a result, there can be radical differences in the accounts and predictions of scholars working in the same places whose work is separated by two or three decades. The concept of "incipient bilingualism" (Diebold 1961) is useful for analyzing language patterns in places with a long history of contact. In such contexts, knowledge of other languages can be gradually accumulated by speakers, setting up "interlingual identifications" that lay the groundwork for rapid future shifts in usage catalyzed by sociocultural change (Diebold 1961: 111). The scholarly assessment of a given speaker's knowledge of particular linguistic varieties in their repertoire can also be fraught with difficulties, as the differences can be vast among what speakers produce in ordinary spoken interactions, how they say they speak in interviews, and what they tacitly know.

For instance, these complexities are evident in a cluster of Tarascan-speaking communities in Mexico that have long been under pressure to shift toward Spanish, the dominant language of political, economic, and educational spheres. While these communities share both geographic proximity and a common sociolinguistic heritage, some are shifting toward Spanish monolingualism while others seem to be maintaining stable bilingualism between Spanish and Tarascan. This demonstrates that "at the cultural level, the speakers in these differentially bilingual

communities are governed by conflicting and complexly ordered values and attitudes" (Friedrich 1971: 168). In other words, even when historical, political, and economic factors are similar, disparate language ideologies can lead to distinct outcomes of language contact. Indeed, among the most crucial factors shaping outcomes of language contact are speakers' social networks and communities, as it is in daily interactions that language shift and maintenance takes place (see Gal 1978; Sicoli 2011). Speakers of endangered languages' social networks are additionally complicated by the recent global increase in transnational migration, which sometimes leads to further language shift in home communities of returning migrants (Perez-Baez 2014).

Among Native American groups throughout the United States, processes of language shift away from the use of native languages toward English are extremely widespread. At present, the vast majority of native languages are considered endangered. This is not a new phenomenon. In fact, language endangerment was a major motivating factor in Franz Boas' focus on "salvage ethnography," in which he and his students worked to document the linguistic and cultural practices of Native American communities before they disappeared. Historical linguistic oppression, such as the forced English language education of Native American youth in boarding schools, continues to impact processes of language socialization in subsequent generations (Webster 2010). These dynamics are at play even among native communities that were at one time held up by scholars as "paragons of persistence" such as the Arizona Tewa (see Kroskrity 2009), instigating widespread concern that Native American youth do not, or will not, have access to the rich poetic and descriptive resources of their heritage tongues (see Kroskrity and Field 2009).

Processes of language shift and endangerment are perpetuated by the marginalization of indigenous languages, particularly those without a written tradition. Such languages are increasingly circumscribed to esoteric domains of use, in contrast with the growing global reach of world languages such as English, which is often considered the dominant language of scientific and technological discourse. Even in communities where members are striving to promote revitalization through the creation of native-language curricula, the ideological dominance of colonial languages can be unintentionally replicated in pedagogical materials (see Meek and Messing 2007). Youth in such contexts face a double bind. They are the hope for the future of language minority communities but also are blamed (or shamed) as the source of cultural and linguistic degeneration (see Bonner 2001; Reynolds 2009; Falconi 2011). Especially where indigenous and European languages coexist within the same or overlapping speech communities and are drawn on strategically by speakers across contexts, scholars grapple alongside community members with how to redefine belonging (see Muehlmann 2008; Kroskrity and Field 2009; Webster 2010; Falconi 2013).

5.2 Language maintenance and revitalization

Whether or not language shift, endangerment, or obsolescence follow in the wake of language contact is meaningfully dependent upon the political, economic, and social contexts in which speakers are socialized and within which they deploy their linguistic resources. A major obstacle to language maintenance or stable bi/multilingualism is a pervasive ideology of monolingualism associated with discourses of nationalism. Such ideologies (discussed earlier) posit language shift as an inevitable outcome of language contact. For example, a twentieth-century policy of Castilianization implemented in rural Mexican schools in indigenous regions led many to confront the encroachment of Spanish into social life outside of the classroom. In one Zapotec-speaking community in Oaxaca, at least, circumscribing Spanish use to classroom contexts has led to successful language maintenance of an indigenous language (Zapotec) as the code of language socialization and family life. This has resulted in stable bilingualism (see Sicoli 2011).

Other examples of language revitalization efforts across geographic and social contexts reveal a range of strategies that have been adopted to foster language maintenance and/or language shift reversal. These strategies have met with varying degrees of success. On Native American reservations in the United States, community-based schooling that implements heritage language curricula has

proven to be effective in reimagining school as a site for Native youth to learn about and develop competency in their heritage languages and cultural traditions (Dick and McCarty 1996). In other cases, community members and social groups have drawn on the power of media and technology to create new forums for using, performing, and engaging with heritage languages (Debenport 2011; Swinehart 2012). Some large-scale approaches to language maintenance have involved political organizing by minority groups to promote multiculturalism and the inclusion of a diverse range of voices as an instantiation of democratic principles (French 2003; Schieffelin and Doucet 1998). Perhaps most exciting and promising is the training of native linguists to write and analyze their own languages (a return to and expansion of methods long practiced by anthropologists), which offers a new array of possibilities for revitalization across global contexts. Linguistic anthropologists in these places and spaces investigate how such developments facilitate new forms of dialogue and, indeed, role reversals between researchers and research participants. This is an emerging focus of current research that is sure to generate productive engaged scholarship in the future.

6 Verbal Art and Performance

Before concluding this chapter, it is relevant to explore the scholarship on verbal art, a research focus that, while marginalized, continues to produce key theoretical insights for ethnographic research of language and culture. The study of verbal art occupies a special place in the history of linguistic anthropology. Drawing from German philology of the nineteenth century, Franz Boas posited that languages in general and folklore in particular played key roles in shaping (or even determining) a people's "mode of thought" (Boas 1901: 6–7; see Bauman and Briggs 2003). Boas asserted that, in the absence of formal teaching or a scientific tradition, tales, myths and legends acted a storehouse for the ideas for the "volk." As mentioned, one of Boas' consequent goals was to collect key folkloristic texts of various Native American groups before those stories and languages ceased to be remembered (e.g., Boas 1894, 1910). Boas worked closely with reliable "native informants" to create written texts that would become integral in shaping anthropological theory (Briggs and Bauman 1999). His notion of analyzing language to gain insight into culturally specific modes of thought was taken up by his student Edward Sapir (Sapir 1949 [1927]). This was foundational to the development of linguistic relativity – the idea that specific aspects of the language one speaks may shape or even determine thought (Gumperz and Levinson 1996; Whorf 1956 [1939]). Such text-based approaches to folklore remained dominant until the development of the ethnography of communication in the late 1960s and 1970s, at which time scholars shifted their attention to the dynamics of performance (Bauman 1975; Hymes 1975). In fact, some formative works of the ethnography of communication and allied developments were actually analyses of verbal art (Jakobson 1960; Hymes 1971; Labov 1972). More recently, key studies of language socialization, language ideologies, and language contact have all drawn from research on verbal art (e.g., Duranti 1994; Kroskrity 2012; Schieffelin 1986; Silverstein 1984; Woolard 1995).

Unfortunately, in contemporary contexts both linguists and anthropologists often marginalize the study of verbal art. Scholars may see poetics as "merely parasitic upon such 'core' areas of linguistics as phonology, syntax, and semantics, or upon such anthropological fields as economy and social organization" (Bauman and Briggs 1990: 59). In reality, though, verbal art (as a type of performance) offers people the possibility of linguistic/ cultural transformation and reproduction (Briggs and Bauman 1992). In performance, reflexive attention paid to language alongside reflexive language itself invites people to "turn, end or reflect back upon themselves, upon the relations, actions, symbols, meanings, codes, roles, statuses, social structures, ethical and legal rules, and other sociocultural components which make up their public 'selves' " (Turner 1987: 24). Social transformation is thus possible, but reflexivity may also lead to reaffirmation of the validity of cultural categories. Verbal art is not parasitic upon grammar or society, but rather plays a central role in the reconstitution and transformation of both language and culture.

6.1 Poetics and performance

Especially during Boas' time, the analysis of poetics – a mainstay of classical European scholarship (e.g., Aristotle 1987, in the fourth century CE) – of verbal patterns of (nonwhite) people living in nonliterate societies was in and of itself a statement of cultural relativity. This was the basis for text-based approaches to verbal art. Boas organized Native American myths and tales using poetic devices such as lines and stanzas, putting him "on the threshold of ethnopoetics" (Hymes 2003). This opens the door to exploration of how poetic devices common in written works (e.g., alliteration, allusion, meter, parallelism) appear in oral traditions, as well as how these distinct oral traditions may utilize conventions not found in European-derived poetic traditions (Giorgio and Giannattasio 2004). Poetics have been discussed in research contexts as diverse as Warao women's ritual wailing in Venezuela, improvisatory poetic dueling in a North Yemeni tribe, and the sonic structuring of sentiment among Kaluli speakers of Papua New Guinea (Briggs 1993; Caton 1990; Feld 1982).

Text-based approaches to verbal art are complemented by performance-based approaches. In contrast to how the term is used in formalist linguistics, here performance is defined as "cultural behavior for which a person assumes responsibility to an audience" (Hymes 1975: 18). Across cultures, speakers use conventionalized poetic devices to invoke performance frames (Bateson 2000 [1972]; Bauman 1975; Goffman 1974). Some of these devices are more explicit than others. For example, "Once upon a time" invokes a story or fairytale; "Knock knock" invokes a specific joking genre; and "Baruch atah Adonai" invokes a Hebrew prayer. Framing one's talk as performance through the use of such devices directs addressees (an audience) to look to the form of the message itself as a conveyor of meaning – in other words, to focus on the poetic function of language (Jakobson 1960). This shift in the co-construction of meaning leads to a focus on distinct forms of nonreferential interpretation, which allows people to address topics or issues otherwise outside the scope of normative conversation (e.g., Abu-Lughod 1985; Desjarlais 1991; Sherzer 2002). The communicative ambiguity that results may mitigate some of aspects of social vulnerability, giving performers and audience members the opportunity to "practice" emotions and perspectives in a low-consequence environment (Beeman 2010).

Increased responsibility to an audience for the content and form of one's words is interconnected with heightened reflexivity in performance. Reflexivity is a term for both phenomenological and linguistic analysis. Here, the capacity of language to refer to itself parallels (or some might argue, produces) the phenomenon of reflexive verbal awareness. Phenomenologically, the term "refers to the capacity of subjects to be aware of themselves *as* subjects and to be aware that they are the focus of another's attention" (Berger and Del Negro 2002: 64). In language, reflexivity is especially found in metapragmatics and reported speech (Lucy 1993). In addition to the use of poetic devices and archaic codes (e.g., the use of Latin in Catholic services), reflexive language patterns such as appeals to tradition or disclaimers of performance may invoke a performance frame (Bauman 1975: 295). For instance, it has been noted that in traditional storytelling Arizona Tewa narrators use the Tewa evidential particle *ba* (translated as "so they say") at least once per clause. This yields "the ego-effacing voice which aligns the storyteller with a traditional past, thus using linguistic practice to locate – or, in this case, to erase – the identity of the speaker" (Kroskrity 2009: 45). In terms of the centrality of verbal art to linguistic and cultural processes, it is significant that this research on Tewa storytelling has been an integral part of long-term research on language maintenance, language shift, and language ideologies (see Kroskrity 2012).

6.2 Intertextuality, circulation, and voice

As the above discussion indicates, grammatical patterning of language at all levels is inter-related with the constitution of culturally specific traditions of verbal art. The formation of all genres, registers, and discourses involves entextualization, a process through which these

grammatical, lexical, and phrasal elements are repeated and become conventional. This leads to the construction metaphorical or literal "texts" – relatively stable phrases or collections of phrases that circulate among speakers (Agha 2005; Bauman 2004). The study of intertextuality and circulation is perhaps the most visible contribution of research on verbal art to linguistic anthropology in general (see Silverstein and Urban 1996). Scholars have worked to understand the constitution and negotiation of social power through analysis of how intertextual gaps – differences between people's conceptualizations of "texts" or genres – are managed in interaction. Particular texts accumulate social value through their connections to other speech events and interactional histories, such that to invoke a text is to harness a source of social power (Falconi 2013; Kulick 1992). At the same time, people who occupy valued social roles may have more authority to utilize such texts (Briggs and Bauman 1992). A simple example of this is the use of the phrase "I now pronounce you…" during marriages. At first glance a performative approach might seem sufficient. That is, these words, articulated by the right person (a priest, a judge) in the right context, seem to perform the social act of wedding (see Austin 1975). However, in actual practice there is considerable variation and play involved in the management of both this simple text and the intertextual gaps that arise through its use. For instance, one might consider legal battles over the validity of gay marriages performed by particular authorities in particular states, or the recontextualization of these words in comedy and in ideological battles (e.g., "I now pronounce you Chuck and Larry"). Viewing such phrases as "texts" allows scholars to trace their circulation through different social realms and evaluate the significance of recontextualization. This approach is especially useful for understanding the circulation of texts in our contemporary media-saturated world (e.g., Thurlow and Mroczek 2011).

Viewed from the perspective of speaker rather than text, such an understanding of circulation means that verbal artists (and speakers more generally) are constantly incorporating past speakers' voices into current performances. In one sense, all utterances, novel or not, are constructed from the multiple voices of people who have spoken in the past (Bakhtin 1981, 1986). In verbal art, reflexive attention paid to the poetic function of language intensifies this process of heteroglossia. In performance, people often incorporate of the voices of contemporary interlocutors or ancestors. This may diffuse the moral weight of responsibility ordinarily placed on a speaker (Irvine 1992; Black 2013). Furthermore, drawing from conventionalized texts, registers, or particular language varieties may allow a performer to represent and synthesize distinct ideas or ideologies in novel ways (Hill 1995). For many scholars of verbal art, the concept of voice thus places the grammatical structuring of poetics and the circulation of text back within its proper place – the lived context of embodied human experiences of identity, power, and social differentiation (Feld et al. 2004: 341; see also Barthes 1977).

7 Conclusion

In the past fifteen years, scholars have utilized the theories described above and expanded upon them to further explore the integral role of language in the reproduction and transformation of social categories such as race, ethnicity, class, and gender (e.g., Bailey 2000; Bucholtz 2001; Gaudio 2011; Hall 2003). Much work on these topics asserts that such social categories are intersectional – that is, they are meaningfully interdependent on one another and mutually reinforced – and they are always constituted in interaction (Alim and Reyes 2011). Thus they are, in theory, subject to confrontation and alteration. This and other contemporary research continues the long tradition in ethnolinguistics and linguistic anthropology of dialogue with public discourses and debates, a practice now referred to as engaged anthropology (Low and Merry 2010). While engagement is touted as a new trend in the social sciences, as far back as the early 1900s Franz Boas and his students were doing work in dialogue with public discourses of their time – for instance, documenting and preserving endangered languages and confronting dominant stereotypes that

essentialized race, language, and culture (Boas 1982). Scholarship on language endangerment, multilingualism, and language in education has long demonstrated the relevance of linguistic analysis for understanding the complexities of the reproduction of inequality and advocated on behalf of marginalized communities. Such research is complemented by ongoing work on activism and social justice (e.g., Jaffe 2012; Zentella 1997). The core theoretical perspectives explored above, including work on language socialization, language ideologies, language contact, and verbal art, provide a formidable toolkit for conducting engaged research and for analyzing the place of language in social life across cultures.

RELEVANT JOURNALS AND SOCIETIES

Journals
American Anthropologist
American Ethnologist
Annual Review of Anthropology
Anthropological Linguistics
Anthropological Quarterly
Discourse and Society
Discourse Studies
Journal of American Folklore
Journal of Linguistic Anthropology
Journal of Pragmatics
Language in Society
Language and Communication
Language and Education
Mind, Culture and Activity
Narrative Inquiry
Pragmatics
Pragmatics and Society

Societies
American Anthropology Association
Australian Anthropological Society
International Pragmatics Association
Society for Linguistic Anthropology

EMERGING TRENDS AND RESEARCH QUESTIONS

1 Language and experience: A growing number of researchers are examining the connections between language and experience, juxtaposing research in linguistic and psychological anthropology. A key resource for this synthesis is phenomenological work on the concept of intersubjectivity.
2 Materiality: Research on materiality focuses on how language is embedded in the material world as well as how linguistic practices are integral in the constitution of the material conditions in which communities and individuals live. One point of discussion among scholars studying materiality is how language intersects with neoliberalism and other socioeconomic patterns on a global scale.
3 Racialization, inequality, and embodiment: Building on the idea that social categories such as class, ethnicity, and even race are constituted in interaction, recent work discusses how the body, and especially particular parts of the body, become saturated with semiotic value and thus integral to processes of racialization and reproduction of inequality.
4 Scale-making: The concept of scale is an alternative to the notions of microanalysis versus macroanalysis. In research on scale-making, scholars discuss that ways that speakers actively draw from linguistic resources across multiple scales and create ideological associations that (re)constitute socioeconomic distinctions and relationships of inequality.

NOTE

1 We discuss these theories because they provide an introduction to research topics discussed throughout the chapter. Another key theory not discussed is participation (see Philips 1972; Goffman 1981; Goodwin and Goodwin 2004). In addition, there are many significant influences on this and other work, including but not limited to speech act theory (Austin 1975; Searle 1969), Wittgenstein's language philosophy (Wittgenstein 1960 [1958]), and borrowings from the Prague school of linguistics (Trubetzkoy 1969 [1939]) and Russian literary theory (Bakhtin 1981).

REFERENCES

Abu-Lughod, L. (1985). Honor and sentiments of loss in a Bedouin society. *American Ethnologist*, 12, 2, 245–61.

Abu-Lughod, L. (1991). Writing against culture. In R. Fox (ed.), *Recapturing Anthropology* (pp. 137–62). Santa Fe, NM: School of American Research.

Agha, A. (2005). Voice, footing, enregisterment. *Journal of Linguistic Anthropology*, 15, 1, 38–59.

Ahearn, L. (2011). *Living Language: An Introduction to Linguistic Anthropology*. Malden, MA: Wiley Blackwell.

Alim, H. S. and Reyes, A. (eds) (2011). Complicating race: articulating race across multiple social dimensions. Special issue of *Discourse and Society*, 22, 4.

Alim, H. S. and Smitherman, G. (2012). *Articulate While Black: Barack Obama, Language, and Race in the US*. New York: Oxford University Press.

Aristotle (1987). *Poetics*, translated by R. Janko. Indianapolis, IN: Hackett.

Austin, J. L. (1975). *How to Do Things with Words*. Cambridge, MA: Harvard University Press.

Bernard, H. R. (2011). *Research Methods in Anthropology: Qualitative and Quantitative Methods*. New York: Alta Mira.

Bailey, B. (2000). The language of multiple identities among Dominican Americans. *Journal of Linguistic Anthropology*, 10, 2, 190–223.

Bakhtin, M. (1981). *The Dialogic Imagination: Four Essays*, translated by C. Emerson. Austin, TX: University of Texas Press.

Bakhtin, M. (1986). The problem of speech genre. In C. Emerson and M. Holquist (trans.), *Speech Genres and Other Essays* (pp. 60–102). Austin, TX: University of Texas Press.

Barthes, R. (1977). The grain of the voice. *Image, Music, Text* (pp. 179–89). London: Fontana.

Basso, K. H. (1988). "Speaking with names": language and landscape among the Western Apache. *Cultural Anthropology*, 3, 2, 99–130.

Bateson, G. (2000 [1972]). A theory of play and fantasy. In G. Bateson, *Steps to an Ecology of Mind* (pp. 177–93). Chicago, IL: University of Chicago Press.

Bauman, R. (1975). Verbal art as performance. *American Anthropologist*, 77, 2, 290–311.

Bauman, R. (2004). *A World of Others' Words: Cross-Cultural Perspectives on Intertextuality*. Malden, MA: Blackwell.

Bauman, R. and Briggs, C. L. (1990). Poetics and performance as critical perspectives on language and social life. *Annual Review of Anthropology*, 19, 59–88.

Bauman, R. and Briggs, C. L. (2003). *Voices of Modernity: Language Ideologies and the Production of Inequality*. Cambridge: Cambridge University Press.

Beeman, W. O. (2010). Performance pragmatics, neuroscience and evolution. *Pragmatics and Society*, 1, 1, 118–37.

Berger, H. and Del Negro, G. (2002). Bauman's *Verbal Art* and the social organization of attention: the role of reflexivity in the aesthetics of performance. *Journal of American Folklore*, 115, 455, 62–91.

Black, S. P. (2013). Narrating fragile stories About HIV/AIDS in South Africa. *Pragmatics and Society*, 4, 3, 345–68.

Boas, F. (1894). *Chinook Texts*. Washington, DC: Smithsonian Institution.

Boas, F. (1901). The mind of primitive man. *Journal of American Folklore*, 14, 52, 1–11.

Boas, F. (1910). *Kwakiutl Tales*. New York: Columbia University Press.

Boas, F. (1911). Introduction. In F. Boas (ed.), *Handbook of North American Indian Languages, Part 1* (pp. 1–83). Washington, DC: US Government Printing Office.

Boas, F. (1982). *Race, Language and Culture*. Chicago, IL: University of Chicago Press.

Bonner, D. M. (2001). Garifuna children's language shame: ethnic stereotypes, national affiliation,

and transnational immigration as factors in language choice in southern Belize. *Language in Society*, 30, 1, 81–96.
Bourdieu, P. (1977) *Outline of a Theory of Practice*. Cambridge: Cambridge University Press.
Bourdieu, P. (1991) *Language and Symbolic Power*. Malden, MA: Polity.
Briggs, C. L. (1993). Personal sentiments and polyphonic voices in Warao women's ritual wailing: music and poetics in critical and collective discourse. *American Anthropologist*, 95, 4, 929–57.
Briggs, C. L. and Bauman, R. (1992). Genre, intertextuality, and social power. *Journal of Linguistic Anthropology*, 2, 2, 131–72.
Briggs, C. L. and Bauman, R. (1999). "The foundation of all future researches": Franz Boas, George Hunt, Native American texts, and the construction of modernity. *American Quarterly*, 51, 3, 479–528.
Brown, P. (2012). The cultural organization of attention. In A. Duranti, E. Ochs, and B. Schieffelin (eds), *The Handbook of Language Socialization* (pp. 29–55). Malden, MA: Wiley Blackwell.
Bucholtz, M. (1999). "Why be normal?": language and identity practices in a community of nerd girls. *Language in Society*, 28, 2, 208–23.
Bucholtz, M. (2000). The politics of transcription. *Journal of Pragmatics*, 32, 10, 1439–65.
Bucholtz, M. (2001). The whiteness of nerds: superstandard English and racial markedness. *Journal of Linguistic Anthropology*, 11, 1, 84–100.
Cameron, D. (1995). *Verbal Hygiene: The Politics of Language*. New York: Routledge.
Caton, S. C. (1990). *"Peaks of Yemen I Summon": Poetry as Cultural Practice in a North Yemeni Tribe*. Los Angeles, CA: University of California Press.
Chomsky, N. (1959). Review of *Verbal Behavior* by B. F. Skinner. *Language*, 35, 26–58.
Chomsky, N. (1964). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
Clifford, J. and Marcus, G. (eds) (1986). *Writing Culture: The Poetics and Politics of Ethnography*. Berkeley, CA: University of California Press.
Conley, J. and O'Barr, W. (1990). *Rules Versus Relationships: The Ethnography of Legal Discourse*. Chicago, IL: University of Chicago Press.
Debenport, E. (2011). As the rez turns: anomalies within and beyond the boundaries of a pueblo community. *American Indian Culture and Research Journal*, 35, 2, 87–110.
Desjarlais, R. R. (1991). Poetic transformation of Yolmo "sadness." *Culture, Medicine, and Psychiatry*, 15, 387–420.
Dick, G. S. and McCarty, T. L. (1996). Reclaiming Navajo: language renewal in an American Indian community school. In N. H. Hornberger (ed.), *Indigenous Literacies in the Americas: Language Planning from the Bottom Up* (pp. 69–94). New York: Mouton de Gruyter.
Diebold, R. (1961). Incipient bilingualism. *Language*, 37, 1, 97–112.
Dorian, N. (1992). *Investigating Obsolescence: Studies in Language Contraction and Death*. Cambridge: Cambridge University Press.
Duranti, A. (1990). Politics and grammar: agency in Samoan political discourse. *American Ethnologist*, 17, 4, 646–66.
Duranti, A. (1994). *From Grammar to Politics: Linguistic Anthropology in a Western Samoan Village*. Los Angeles, CA: University of California Press.
Duranti, A. (1997). *Linguistic Anthropology*. Cambridge: Cambridge University Press.
Duranti, A. (2003). Language as culture in US anthropology: three paradigms. *Current Anthropology*, 44, 3, 323–47.
Duranti, A. (2006). Transcripts, like shadows on a wall. *Mind, Culture, and Activity*, 13, 4, 301–10.
Duranti, A. and Black, S. P. (2012). Language socialization and verbal improvisation. In A. Duranti, E. Ochs, and B. Schieffelin (eds), *The Handbook of Language Socialization* (pp. 443–63). Malden, MA: Wiley Blackwell.
Enfield, N. J., Kockelman, P., and Sidnell, J. (eds) (2014). *The Cambridge Handbook of Linguistic Anthropology*. Cambridge: Cambridge University Press.
Fader, A. (2009). *Mitzvah Girls: Bringing Up the Next Generation of Hasidic Jews in Brooklyn*. Princeton, NJ: Princeton University Press.
Falconi, E. (2011). *Migrant Stories: Zapotec Transborder Migration and the Production of a Narrated Community*. PhD dissertation, University of Michigan, Ann Arbor.
Falconi, E. (2013). Storytelling, language shift, and revitalization in a transborder community: "Tell it in Zapotec!" *American Anthropologist*, 115, 4, 622–36.
Feld, S. (1982). *Sound and Sentiment: Birds, Weeping, Poetics, and Song in Kaluli Expression*. Philadelphia, PA: University of Pennsylvania Press.
Feld, S., Fox, A., Porcello, T., and Samuels, D. (2004). Vocal anthropology: from the music of language to the language of song. In A. Duranti (ed.), *A Companion to Linguistic Anthropology* (pp. 321–46). Malden, MA: Blackwell.
Finlayson, R. (2002). Women's language of respect: *Isihlonipho Sabafazi*. In R. Mesthrie (ed.), *Language in South Africa* (pp. 279–96). New York: Cambridge University Press.
Foucault, M. (1972). *The Archaeology of Knowledge and The Discourse on Language*. New York: Pantheon Books.
Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings, 1972–77*. Random House.

French, B. (2003). The politics of Mayan linguists in Guatemala: native speakers, expert analysts and the nation. *Pragmatics*, 13, 4, 483–98.

Friedrich, P. (1971). Dialectal variation in Tarascan phonology. *International Journal of American Linguistics*, 37, 3, 164–87.

Gal, S. (1978). Peasant men can't get wives: language change and sex roles in a bilingual community. *Language in Society*, 7, 1, 1–16.

Gal, S. (1987). Code switching and consciousness on the European periphery. *American Ethnologist*, 14, 637–53.

Gal, S. (1998) Multiplicity and contention among language ideologies: a commentary. In B. Schieffelin, K. Woolard, and P. V. Kroskrity (eds), *Language Ideology: Practice and Theory* (pp. 317–31). Oxford: Oxford University Press.

Gal, S. (2012). The role of language in the ethnographic method. In R. Fardon, O. Harris, T. H. J. Marchland, M. Nuttall, C. Shore, … R. A. Wilson (eds), *The Sage Handbook of Social Anthropology* (pp. 38–53). Los Angeles, CA: Sage.

García-Sánchez, I. M. (2014). *Language and Muslim Immigrant Childhoods: The Politics of Belonging*. Oxford: Wiley Blackwell.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Malden, MA: Blackwell.

Garrett, P. (2004). Language contact and contact languages. In A. Duranti (ed.), *A Companion to Linguistic Anthropology*. Malden, MA: Wiley Blackwell.

Garrett, P. (2005). What a language is good for: language socialization, language shift, and the persistence of code-specific genres in St. Lucia. *Language in Society*, 34, 3, 327–61.

Garrett, P. and P. Baquedano-Lopez (2002). Language socialization: reproduction and continuity, transformation and change. *Annual Review of Anthropology*, 31, 339–61.

Gaudio, R. (2011). The blackness of "broken English." *Journal of Linguistic Anthropology*, 21, 2, 230–46.

Geertz, C. (1983). From the native's point of view: on the nature of anthropological understanding. In C. Geertz, *Local Knowledge: Further Essays in Interpretive Anthropology* (pp. 55–70). New York: Basic Books.

Giddens, A. (1979). *Central Problems in Social Theory: Action, Structure, and Contradiction in Social Analysis*. Berkeley, CA: University of California Press.

Giorgio, B. and Giannattasio, F. (2004). Poetry. In A. Duranti (ed.), *A Companion to Linguistic Anthropology* (pp. 290–320). Malden, MA: Blackwell.

Goffman, E. (1974). *Frame Analysis: An Essay on the Organization of Experience*. Boston, MA: Northeastern University Press.

Goffman, E. (1981). *Forms of Talk*. Philadelphia, PA: University of Pennsylvania Press.

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (ed.), *Everyday Language: Studies in Ethnomethodology* (pp. 97–121). New York: Irvington.

Goodwin, C. (1981). *Conversational Organization: Interaction Between Speakers and Hearers*. New York: Academic Press.

Goodwin, C. (2004). A competent speaker who can't speak: the social life of aphasia. *Journal of Linguistic Anthropology*, 14, 2, 151–70.

Goodwin, C. and Goodwin, M. (2004). Participation. In A. Duranti (ed.), *A Companion to Linguistic Anthropology* (pp. 222–44). Malden, MA: Wiley Blackwell.

Goodwin, M. and Goodwin, C. (2000). Emotions within situated activity. In N. Budwig, I. C. Uzigiris, and J. V. Wertsch (eds), *Communication: An Arena of Development* (pp. 33–53). Westport, CT: Greenwood Publishing.

Gramsci, A. (1971). *Selections from The Prison Notebooks*, edited by Q. Hoare and G. N. Smith. New York: International Publishers.

Gumperz, J. (1964). Linguistic and social interaction in two communities. *American Anthropologist*, 66, 6–2, 137–53.

Gumperz, J. (1982). *Discourse Strategies*. Cambridge: Cambridge University Press.

Gumperz, J. and Hymes, D. (eds) (1964). The ethnography of communication. *American Anthropologist*, 66, 6–2.

Gumperz, J. and Levinson, S. (eds) (1996). *Rethinking Linguistic Relativity*. New York: Cambridge University Press.

Hall, K. (2003) Exceptional speakers: contested and problematized gender identities. In M. Meyerhoff and J. Holmes (eds), *Handbook of Language and Gender* (pp. 352–80). Oxford: Basil Blackwell.

Haviland, J. B. (2004). Gesture. In A. Duranti (ed.), *A Companion to Linguistic Anthropology* (pp. 197–221). Malden, MA: Wiley Blackwell.

Haviland, J. B. (2011). Nouns, verbs, and constituents in an emerging "Tzotzil" sign language. In R. Gutiérrez-Bravo, L. Mikkelsen, and E. Potsdam (eds), *Representing Language: Essays in Honor of Judith Aissen*. Santa Cruz, CA: Linguistics Research Center, UCSC.

Herbert, R. K. (1990a). The sociohistory of clicks in Southern Bantu. *Anthropological Linguistics*, 32, 295–315.

Herbert, R. K (1990b). The relative markedness of click sounds: evidence from language change, acquisition, and avoidance. *Anthropological Linguistics*, 32, 1–2, 120–38.

Heritage, J. (1984). *Garfi nkel and Ethnomethodology*. Cambridge: Polity.

Hill, J. H. (1995). The voices of Don Gabriel: responsibility and self in a modern Mexicano narrative.

In D. Tedlock and B. Mannheim (eds), *The Dialogic Emergence of Culture* (pp. 97–147). Chicago, IL: University of Illinois Press.

Hill, J. H. (1998). Language, race, and white public space. *American Anthropologist*, 100, 3, 680–9.

Hill, J. H. and Hill, K. (1986). *Speaking Mexicano: Dynamics of Syncretic Language in Central Mexico*. Tucson, AZ: University of Arizona Press.

Hinton, L. (2012). *Bringing Our Languages Home: Language Revitalization for Families*. Berkeley, CA: Heyday Press.

Hymes, D. (1963). Notes toward a history of linguistic anthropology. *Anthropological Linguistics*, 5, 1, 59–103.

Hymes, D. (1971). The contribution of folklore to sociolinguistics. *Journal of American Folklore*, 84, 331, 42–50.

Hymes, D. (1972). On communicative competence. In J. B. Pride and J. Holmes (eds), *Sociolinguistics* (pp. 269–93). Harmondsworth: Penguin.

Hymes, D. (1975). Breakthrough into performance. In D. Ben-Amos and K. S. Goldstein (eds), *Folklore: Performance and Communication*, (pp. 11–74). The Hague: Mouton.

Hymes, D. (2003). Franz Boas on the threshold of ethnopoetics. In D. Hymes, *Now I Know Only So Far: Essays in Ethnopoetics* (pp. 15–35). Lincoln, NE: University of Nebraska Press.

Irvine, J. (1979). Formality and informality in communicative events. *American Anthropologist*, 81, 4, 773–90.

Irvine, J. (1989). When talk isn't cheap: language and political economy. *American Ethnologist*, 16, 2, 248–67.

Irvine, J. (1992). Insult and responsibility: verbal abuse in a Wolof village. In J. H. Hill and J. Irvine (eds), *Responsibility and Evidence in Oral Discourse* (pp. 105–34). New York: Cambridge University Press.

Irvine, J. (1998). Ideologies of honorific language. In B. B. Schieffelin, K. Woolard, and P. Kroskrity (eds), *Language Ideologies: Practice and Theory* (pp. 51–67). New York: Oxford University Press.

Irvine, J. (2006). Speech and language community. In K. Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edn (pp. 689–98). Oxford: Elsevier.

Irvine, J. (2012). Language ideology. In Oxford Bibliographies. Oxford: Oxford University Press. Available online at http://www.oxfordbibliographies.com/view/document/obo-9780199766567/obo-9780199766567-0012.xml

Irvine, J. and Gal, S. (2000). Language ideology and linguistic differentiation. In P. Kroskrity (ed.), *Regimes of Language: Ideologies, Polities, and Identities* (pp. 35–83). Santa Fe, NM: School of American Research.

Jackson, M. (1998). *Minima Ethnographica: Intersubjectivity and the Anthropological Project*. Chicago, IL: University of Chicago Press.

Jaffe, A. (1999). *Ideologies in Action: Language Politics on Corsica*. Berlin: De Gruyter Press.

Jaffe, A. (2012). Collaborative practice, linguistic anthropological enquiry and the mediation between researcher and practitioner discourses. In S. Gardner and M. Martin-Jones (eds), *Multilingualism, Discourse and Ethnography* (pp. 334–52). New York: Routledge.

Jakobson, R. (1960). Concluding statement: linguistics and poetics. In T. A. Sebeok (ed.), *Style and Language* (pp. 350–77). Cambridge, MA: MIT Press.

Keane, W. (2007). *Christian Moderns: Freedom and Fetish in the Mission Encounter*. Los Angeles, CA: University of California Press.

Kroskrity, P. (2004). Language ideologies. In A. Duranti (ed.), *A Companion to Linguistic Anthropology* (pp. 496–517). Malden, MA: Wiley Blackwell.

Kroskrity, P. (2009). Narrative reproductions: ideologies of storytelling, authoritative words, and generic regimentation in the village of Tewa. *Journal of Linguistic Anthropology*, 19, 1, 40–56.

Kroskrity, P. (ed.) (2012) *Telling Stories in the Face of Danger: Language Renewal in Native American Communities*. Norman, OK: University of Oklahoma Press.

Kroskrity, P. and Field, N. (eds) (2009). *Native American Language Ideologies: Beliefs, Practices, and Struggles in Indian Country*. Tucson, AZ: University of Arizona Press.

Kulick, D. (1992). *Language Shift and Cultural Reproduction: Socialization, Self, and Syncretism in a Papua New Guinean Village*. Cambridge: Cambridge University Press.

Labov, W. (1966). *The Social Stratification of English in New York City*. Arlington, VA: Center for Applied Linguistics.

Labov, W. (1972). *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia, PA: University of Pennsylvania Press.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press.

Lippi-Green, R. (1997). *English with an Accent: Language, Ideology, and Discrimination in the United States*. New York: Routledge.

Low, S. M. and Merry, S. E. (2010). Engaged anthropology: diversity and dilemmas: an introduction to supplement 2. *Current Anthropology*, 51, S2, S203–26.

Lucy, J. (ed.) (1993). *Reflexive Language: Reported Speech and Metapragmatics*. New York: Cambridge University Press.

Malinowski, B. (1935). The language of magic and gardening. In B. Malinowski, *Coral Gardens and*

Their Magic: A Study of the Methods of Tilling the Soil and Agricultural Rites in the Trobriand Islands, vol. 2. New York: American Book Company.

Mannheim, Bruce. (1984). Una nación acorralada: Southern Peruvian Quechua language planning and politics in historical perspective. *Language in Society*, 13, 3, 291–309.

Marcus, G. (2008). The ends of ethnography: social/cultural anthropology's signature form of producing knowledge in transition. *Cultural Anthropology*, 23, 1, 1–14.

Meek, B. (2010). *We Are Our Language: An Ethnography of Language Revitalization in a Northern Athabaskan Community*. Tucson, AZ: University of Arizona Press.

Meek, B. and Messing, J. (2007). Framing indigenous languages as secondary to matrix languages. *Anthropology and Education Quarterly*, 38, 2, 99–118.

Milroy, L. (2001) Britain and the United States: two nations divided by the same language (and different language ideologies). *Journal of Linguistic Anthropology*, 10, 1, 56–89.

Morgan, M. (2004). Speech community. In A. Duranti (ed.), *A Companion to Linguistic Anthropology* (pp. 3–22). Malden, MA: Wiley Blackwell.

Muehlman, S. (2008). "spread your ass cheeks" and other things that should not be said in indigenous languages. *American Ethnologist*, 35, 1, 34–48.

Ochs, E. (1979). Transcription as theory. In E. Ochs and B. B. Schieffelin (eds), *Developmental Pragmatics* (pp. 43–72). New York: Academic Press.

Ochs, E. (1982). Ergativity and word order in Samoan child language. *Language*, 58, 646–71.

Ochs, E. (1992). Indexing gender. In A. Duranti and C. Goodwin (eds), *Rethinking Context: Language as an Interactive Phenomenon* (pp. 335–58). New York: Cambridge University Press.

Ochs, E. and Schieffelin, B. (1984). Language acquisition and socialization: three developmental stories and their implications. In R. Shweder and R. LeVine (eds), *Culture Theory: Essays on Mind, Self and Emotion* (pp. 276–320). New York: Cambridge University Press.

Peirce, C. (1985 [1897]). Logic as semiotic: the theory of signs. In R. Innis (ed.), *Semiotics: An Introductory Anthology* (pp. 4–23). Bloomington, IN: Indiana University Press.

Perez-Baez, G. (2014). Determinants of language reproduction and shift in a transnational community. *International Journal of the Sociology of Language*, 227, 65–81.

Philips, S. U. (1972). Participant structures and communicative competence. In C. Cazden, V. John, and D. Hymes (eds), *The Function of Language in the Classroom* (pp. 370–94). New York: Teachers College Press.

Philips, S. U. (1991). Tongan speech levels: practice and talk about practice in the cultural construction of social hierarchy. In Robert Blust (ed.), *Currents in Pacific Linguistics: Papers on Austronesian Languages and Ethnolinguistics in Honour of George W. Grace* (pp. 369–82). Canberra: Pacific Linguistics C–117.

Philips, S. U. (1998). Language ideologies in institutions of power: a commentary. In B. B. Schieffelin, K. Woolard, and P. Kroskrity (eds), *Language Ideologies: Practice and Theory* (pp. 211–25). New York: Oxford University Press.

Philips, S. (2000). Constructing a Tongan nation-state through language ideology in the courtroom. In P. Kroskrity (ed.), *Regimes of Language: Ideologies, Polities, and Identities*. Santa Fe, NM: School of American Research Press.

Philips, S. U. (2013). Method in anthropological discourse analysis: the comparison of units of interaction. *Journal of Linguistic Anthropology*, 23, 1, 82–95.

Phillips, J. (1973) Syntax and vocabulary of mothers' speech to young children: age and sex comparisons. *Child Development*, 44, 182–5.

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: Harper Collins.

Preston, D. (1996). Whaddayaknow?: The modes of folk linguistic awareness. *Language Awareness*, 5, 1, 40–74.

Raymond, G. (2009). Grammar and social relations: grammar as a constitutive feature of conversational and institutional activities. In A. Freed and S. Erlich (eds), *"Why Do You Ask?" The Function of Questions in Institutional Discourse* (pp. 87–107). Oxford: Oxford University Press.

Reynolds, J. (2009). Shaming the shift generation: intersecting ideologies of family and linguistic revitalization in Guatemala. In P. Kroskrity and N. Field (eds), *Native American Language Ideologies: Beliefs, Practices, and Struggles in Indian Country*. Tucson, AZ: University of Arizona Press.

Rumsey, A. (1990). Wording, meaning, and linguistic ideology. *American Anthropologist*, 92, 346–61.

Rumsey, A. (2013). Anthropology, linguistics, and the vicissitudes of interdisciplinary collaboration. *Collaborative Anthropologies*, 6, 268–89.

Rymes, B. (1996). Naming as social practice: the case of Little Creeper from Diamond Street. *Language in Society*, 25, 2, 237–60.

Sachs, J., Brown, R. and Salerno, R. (1976). Adults' speech to children. In W. von Raffler Engel and Y. Lebrun (eds), *Baby Talk and Infant Speech*. Lisse: Riddler Press.

Sacks, H. (1984). On doing "being ordinary." In J. M. Atkinson and J. C. Heritage (eds), *Structures of Social Action* (pp. 413–29). Cambridge: Cambridge University Press.

Sacks, H., Schegloff, E. and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 4, 696–735.

Sapir, E. (1949 [1927]). The unconscious patterning of behavior in society. In D. G. Mandelbaum (ed.), *Selected Writings of Edward Sapir in Language, Culture and Personality* (pp. 544–59). Los Angeles, CA: University of California Press.

Saussure, F. de (1959 [1916]). *Course in General Linguistics*, translated by R. Harris. New York: McGraw-Hill.

Schegloff, E. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29, 1, 1–63.

Schieffelin, B. B. (1986). Teasing and shaming in Kaluli children's interactions. In B. B. Schieffelin and E. Ochs (eds), *Language Socialization across Cultures* (pp. 165–81). New York: Cambridge University Press.

Schieffelin, B. B. (1990). *The Give and Take of Everyday Life: Language Socialization of Kaluli Children*. New York: Cambridge University Press.

Schieffelin, B. B. and Cochran-Smith, M. (1984). Learning to read culturally: literacy before schooling. In H. Goelman, A. Oberg, and F. Smith (eds), *Awakening to Literacy*. Exeter, NH: Heinemann.

Schieffelin, B. B. and Doucet, R. (1998). The "real" Haitian creole: ideology, metalinguistics, and orthographic choice. *American Ethnologist*, 21, 1, 176–200.

Schieffelin, B. B. and Ochs, E. (1986). Language socialization. *Annual Review of Anthropology*, 15, 163–91.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Senghas R. J. (1997). An "unspeakable, unwriteable" language: deaf identity, language and personhood among the first cohorts of Nicaraguan signers. PhD thesis, University of Rochester.

Senghas R. J. and Monaghan, L. (2002). Signs of their times: deaf communities and the culture of language. *Annual Review of Anthropology*, 31, 69–97.

Sherzer, J. (2002). *Speech Play and Verbal Art*. Austin, TX: University of Texas Press.

Sicoli, M. (2011). Agency and ideology in language shift and maintenance. In T. Granadillo and H. Orcutt-Gachiri (eds), *Ethnographic Contributions in the Study of Endangered Languages: A Linguistic Anthropological Perspective*. Tucson, AZ: University of Arizona Press.

Sidnell, J. and Stivers, T. (eds) (2012). *The Handbook of Conversation Analysis*. Malden, MA: Wiley Blackwell.

Silverstein, M. (1976). Shifters, linguistic categories, and cultural description. In K. H. Basso and H. A. Selby (eds), *Meaning in Anthropology* (pp. 11–55). Albuquerque, NM: University of New Mexico Press.

Silverstein, M. (1979). Language structure and linguistic ideology. In P. Clyne (ed.), *The Elements: A Parasession on Linguistic Units and Levels* (pp. 193–247). Chicago, IL: Chicago Linguistic Society.

Silverstein, M. (1981). The limits of awareness. *Sociolinguistic Working Paper*, 84. Austin, TX: Southwest Educational Development Laboratory.

Silverstein, M. (1984). On the pragmatic 'poetry' of prose: parallelism, repetition, and cohesive structure in the time course of dyadic conversation. In D. Schiffrin (ed.), *Meaning, Form, and Use in Context: Linguistic Applications* (pp. 181–99). Georgetown: Georgetown University Press.

Silverstein, M. (1985). Language and the culture of gender: at the intersection of structure, usage, and ideology. In E. Mertz and R. J. Parmentier (eds), *Semiotic Mediation* (pp. 219–59). Orlando, FL: Academic Press.

Silverstein, M. (1993). Metapragmatic discourse and metapragmatic function. In J. Lucy (ed.), *Reflexive Language* (pp. 33–58). New York: Cambridge University Press.

Silverstein, M. (2006). How we look from where we stand. *Journal of Linguistic Anthropology*, 16, 2, 269–78.

Silverstein, M., and Urban, G. (eds) (1996). *Natural Histories of Discourse*. Chicago, IL: University of Chicago Press.

Spitulnik, D. (1996). The social circulation of media discourse and the mediation of communities. *Journal of Linguistic Anthropology*, 6, 2, 161–87.

Stocking, G. W., Jr. (1966). Franz Boas and the culture concept in historical perspective. *American Anthropologist*, 68, 4, 867–82.

Streeck, J., Goodwin, C. and LeBaron, C. (eds) (2012). *Embodied Interaction: Language and Body in the Material World*. New York: Cambridge University Press.

Swinehart, K. (2012). Metadiscursive regime and register formation on Aymara radio. *Language and Communication*, 32, 102–13.

Tedlock, D. and Mannheim, B. (eds) (1995). *The Dialogic Emergence of Culure*. Urbana, IL: University of Illinois Press.

Teeter, K. V. (1964). "Anthropological linguistics" and linguistic anthropology. *American Anthropologist*, 66, 4, 878–9.

Thurlow, C. and Mroczek, K. (eds) (2011). *Digital Discourse: Language in the New Media*. New York: Oxford University Press.

Trubetzkoy, N. S. (1969 [1939]). *Principles of Phonology*, translated by C. A. M. Baltaxe. Los Angeles, CA: University of California Press.

Turner, V. (1987). *The Anthropology of Performance*. Baltimore, MD: Johns Hopkins University Press.

Voegelin, C. F. (1965). Sociolinguistics, ethnolinguistics, and anthropological linguistics. *American Anthropologist*, 67, 2, 484–5.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Webb, V. (2002). *Language in South Africa: The Role of Language in National Transformation, Reconstruction, and Development*. Philadelphia, PA: John Benjamins.

Webster, A. (2010). Imagining Navajo in the boarding school: Laura Tohe's *No Parole Today* and the intimacy of language ideologies. *Journal of Linguistic Anthropology*, 20, 1, 39–62.

Whorf, B. L. (1956 [1939]). The relation of habitual thought and behavior to language. In J. B. Carroll (ed.), *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, (pp. 134–59). Cambridge, MA: MIT Press.

Williams, R. (1977). *Marxism and Literature*. Oxford: Oxford University Press.

Wittgenstein, L. (1960 [1958]). *The Blue and Brown Books: Preliminary Studies for the "Philosophical Investigations."* New York: Harper and Row.

Woolard, K. (1995). Changing forms of codeswitching in Catalan comedy. *Catalan Review*, 9, 2, 223–52.

Woolard, K. (1998a). Introduction: Language ideology as a field of inquiry. In B. B. Schieffelin, K. Woolard, and P. Kroskrity (eds), *Language Ideologies: Practice and Theory* (pp. 3–47). New York: Oxford University Press.

Woolard, K. (1998b). Simultaneity and bivalence as strategies in bilingualism. *Journal of Linguistic Anthropology*, 8, 1, 3–29.

Zentella, A. C. (1997). *Growing Up Bilingual: Puerto Rican Children in New York*. Malden, MA: Blackwell.

24 Sociolinguistic Theory

Systematic Study of the Social Uses of Language

J. K. CHAMBERS

"Everyone knows that language is variable," said Edward Sapir (1921: 147). Its variability is indeed a commonplace, almost trivial, observation. It should be surprising, then, to realize that linguistic theories dealing with language as a variable phenomenon are relatively recent. The perception that linguistic variation could be dealt with as a coherent field of study, in other words, as a theory with research issues of international concern, a set of protocols and methods for exploring them, and a cumulative body of empirical results, had its effective beginning only in the second half of the twentieth century. Its time was ripe. The branch of language science is known as sociolinguistics, and it is now firmly established as a core area in the discipline.

1 Concept and Percept

The study of linguistic variation entailed a reconceptualization from the study of language structure for its own sake to the study of language in its many social uses. The two poles, roughly the structural and the social biases, have been characterized in many ways. The essence of language as structure is grammar (Chomsky 1980: 55), the set of rules and representations that determine the combinatorial possibilities that make up the syntax and phonotactics of language. The essence of language as social act is the manipulation of linguistic elements for communicative purposes. The distinction between grammar and utterance belongs to the philosophical antinomy between the ideal and real, shadow and substance, Humboldt's *energeia* and *ergon*, Saussure's *langue* and *parole*, Chomsky's competence and performance (discussed in Chambers 2013a: esp. 6–10).

The philosopher William James distinguished concepts, the idealization of reality, from percepts, the apprehension of reality. "There must always be a discrepancy between concepts and reality," he said (1911: 365), "because the former are static and discontinuous, and the latter are dynamic and flowing." James recognized the primacy of percepts in human experience (1911: 96):

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

> The deeper features of reality are found only in perceptual experience. Here alone do we acquaint ourselves with continuity, or the immersion of one thing in another, here alone with self, with substance, with qualities, with activity in its various modes, with time, with cause, with changes, with novelty, with tendency and with freedom.

James, in common with all thinkers of his generation, believed that the inherent flux in percepts made them unmanageable as objects of study. They could only be apprehended in terms of the concepts of which they were imperfect realizations.

More familiar to linguists is the distinction made by James's contemporary, Ferdinand de Saussure, the founder of modern linguistics. Saussure referred to language-as-concept by the term *langue* (French for "language"), and to language-as-percept as *parole* (French for "word"). Saussure noted that "speech has both an individual and a social side, and we cannot conceive of one without the other" (1916: 8). But he had no doubts about which of the two "sides" of language was worthy of study. *Langue* "is a self-contained whole," he said (1916: 9). "As soon as we give *langue* first place among the facts of speech, we introduce a natural order into a mass that lends itself to no other classification." By contrast, *parole* "is many-sided and heterogeneous; straddling several areas simultaneously – physical, physiological and psychological – it belongs to both the individual and to society; we cannot put it into any category of human facts, for we cannot discover its unity."

Five decades after Saussure, Noam Chomsky made the same distinction in different terms. He spoke of *competence*, "the speaker-hearer's knowledge of his language," as distinguished from *performance*, "the actual use of language in concrete situations" (1965: 4). And, also like Saussure, he had no doubts that competence, not performance, was the proper domain of linguistic study. He said, "The observed use of language . . . surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline."

Throughout the history of language study prior to the emergence of sociolinguistics, numerous linguists had stated explicitly in their different ways that the proper object of study was *langue* or *competence*. Long before Saussure, in the 1830s, Humboldt distinguished language as "an *ergon*, that is a mere means of exchange for purposes of communication," and "an *energeia* which reconstitutes human experience ideally" (Basilius 1952: 98). Joos (1950: 701–2) put it this way: "We must make our 'linguistics' a kind of mathematics within which inconsistency is by definition impossible. . . . All continuities, all possibilities of infinitesimal gradation, are shoved outside of linguistics in one direction or another." Hjelmslev (1961: 5–6) said: "Linguistics must attempt to grasp language, not as a conglomerate of non-linguistic (e.g. physical, physiological, logical, sociological) phenomena, but as a self-sufficient totality, a structure *sui generis*." These statements and countless others like them, disparate though they may seem, are all versions of the Axiom of Categoricity (Chambers 2009: 12, 26–8), the idealization that removes language from the variablity of *parole* or *performance*.

2 *The Science of* Parole

The emergence of sociolinguistics in the 1960s thus ran counter to centuries of precedent in language study from classical Greek grammatical speculation (Chambers 2009: 240–1) to Chomskyan generativism. Sociolinguistics is the science of *parole* or *ergon* or performance. Sociolinguistics seeks to understand language amidst the Hjelmslevian conglomerate of physical, physiological, logical, and sociological phenomena, suffused with all the Joosian continuities, all possibilities of infinitesimal gradation, and (more humbly) precisely as the Humboldtian means of exchange for purposes of communication.

Sociolinguistics aims to come to grips with language, in the words of one of its seminal documents (Weinreich, Labov, and Herzog 1968: 100), "as an object possessing orderly heterogeneity."

There were precedents for sociolinguistics in the long history of systematic language study, but they proved to be isolated insights rather than foundational. The Roman polymath Varro (116–27 BCE) recognized linguistic variation (*anomalia*) and linked it to vernacular language use. The surviving fragment of his treatise on language captures what would become the essence of sociolinguistics in his maxim *consuetudo loquendi est in motu* – "the vernacular is always in motion" (Chambers 2009: 241). Unfortunately, Varro's theories gave rise to no school of thought.

Centuries later, the Swiss dialectologist Louis Gauchat (1866–1942) undertook field research in the alpine village Charmey and then wrote about "speech differences within a single dialect," enumerating "variation in the pronunciation, morphology, syntax, and lexicon of the dialect of a village" (2008 [1905]: 186). "The issue of dialects has been the subject of lively scholarly debate," Gauchat says in his introductory section, "but variation in pronunciation among members of a single speech community has not been studied systematically" (p. 175). He was very self-conscious about the precedent he was setting with his study of variation, even apologetic: "My reason . . . for making public these opinions I have formed on the unity of speech in a single community is to encourage dialectologists to undertake similar research in other places." But it was not to be. Gauchat's approach would not gain adherents for six more decades when sociolinguistics came into being, and by then his precedent was largely forgotten (Chambers 2008).

2.1 The variable as structural unit

Linguists and others ignored variation in language for centuries because they were convinced that it was chaotic, incoherent, and lawless. The first challenge for sociolinguistics was to gather evidence of variability on a large scale and demonstrate that its occurrence in the speech community was systematic, patterned, and orderly. In this respect, sociolinguistics was participating in a revolution in intellectual history that affected several other formerly categorical disciplines in the mid-twentieth century such as mathematics, which admitted fuzzy sets and chaos theory; symbolic logic, which admitted multidimensional truth tables; and statistics, which admitted multivariate analyses. What all these variationist theories share is the admission of categories with intermediate values, between "true" and "false," or "on" and "off," or 1 and 0.

In order to demonstrate how the analysis of variable linguistic data proceeds and what its correlations with independent variables reveal, I will discuss a well-studied variable of English often symbolized (C$\underline{\text{C}}$) but sometimes called (more descriptively) final stop deletion or morpheme-final consonant cluster simplification. The parentheses in (C$\underline{\text{C}}$) characterize a linguistic variable (as slashes characterize phonemes and square brackets phones). Variables are abstract entities that are actualized in utterances by their variants, of which there must be two or more. (A "variable" with a single variant is categorical, and not variable at all.) The Cs in (C$\underline{\text{C}}$) stand for "consonant" (as they do generally in phonology), but in this instance they stand for morpheme-final consonant clusters, in words like *last*, *land*, and many others. The second C in (C$\underline{\text{C}}$) is underlined because stops in this position, for instance, the /t/ in *last* and the /d/ in *land*, are sometimes not pronounced under certain well-defined conditions. In other words, pronouncing the final stop in the cluster is variable, and the variation is conditioned, not random or capricious.

The conditions under which the final consonant may be deleted are both linguistic and social. The linguistic conditions are determined by the phonological elements on both sides of the deletable stop ($\underline{\text{C}}$), as shown in Table 24.1. The first C in the CC cluster can be either a sonorant (Son) like the /n/ in *land*, or an obstruent (Obs) like the /s/ in *last*. The elements following the deletable stop across the word boundary (#) can be a consonant (C) like the /f/ in *land fall* and *last fall*, or a vowel (V) like the /o/ in *land owner* or *last owner*, or a pause (##) when the item occurs in utterance-final position.

Table 24.1 Phonological environments in which the deletable consonant (C) occurs in utterances, either preceded by sonorant (Son) or obstruent (Obs) and followed across the word or morpheme boundary (#) by a consonant (C), vowel (V), or pause (##).

C	PRECEDED BY SONORANT	PRECEDED BY OBSTRUENT
followed by Consonant	...SonC # C... *land fall*	...ObsC # C... *last fall*
followed by Vowel	...SonC # V... *land owner*	...ObsC # V... *last owner*
followed by Pause	...SonC # # *land.*	...ObsC # # *last.*

Simplification of the word-final cluster, that is, the pronunciation of *last* as *las'* or *land* as *lan'*, is not free (or arbitrary) but occurs in well-defined conditions. The constraints on variability for (CC) can be simply demonstrated for native speakers of any standard variety of English (that is, for middle-class speakers of American English, Scottish English, English English, Australian English, Canadian English, and other English varieties), simply by reading the word-list in Table 24.1 with the deletable stop (the /d/ and the /t/) deleted. Perceptually, it will be clear that some of the utterances sound normal with the stop omitted and others sound odd. The normal ones (for all speakers of standard English varieties) are *lan' fall* and *las' fall*; the odd-sounding ones are all the others, that is, *lan' owner*, *lan'*, *las' owner*, and *las'*.

Linguistically, (CC) simplification occurs in standard English varieties only when the deletable C comes before another consonant in the utterance (*lan' fall* and *las' fall*). Graphically, in Table 24.1 standard varieties draw the line at the first row; (CC) simplification does not normally occur lower down. In more general phonological terms, the condition under which the stop can be deleted is when it occurs in a triconsonantal sequence, schematically . . .CC#C. Phonotactically, the effect of (CC) simplification, and its apparent motivation, is to reduce the articulatory gestures required by triconsonantal clusters by realizing them as biconsonantal (CØ#C).

The linguistic condition is necessary but not sufficient in accounting for the probability of simplification. Certain nonlinguistic conditions also play a role. Social class, for one, is crucially involved. The constraints we are talking about, as we have said, hold only for standard varieties, that is, for the English characteristically spoken by middle-class people. Style is also involved. Simplification occurs more frequently in casual speech, that is, in informal situations when friends and acquaintances are speaking to each other.

These constraints on (CC) – that it occurs in middle-class speech mainly before another consonant, and more frequently in casual contexts – were determined by the analysis of hours of natural speech by women and men in numerous communities. They are hardly surprising. With hindsight, they coincide with the introspections of linguistically sensitive people. They could not, however, have been discovered by introspection, at least not reliably. Most constraints appear to go unperceived by even the most sensitive speakers, and many are so subtle and subconscious that they are scarcely accessible to our intuitions even when they are robust in the data. A telling example arose in (CC) research, when it was discovered that frequency of simplification is greater when the preceding C is a sonorant than when it is an obstruent. In other words, *land fall* is more likely to be uttered as *lan' fall* than is *last fall* to be *las' fall* when the other conditions are fulfilled.

So far, we have considered (CC) in a fairly narrow social stratum, namely, middle-class standard varieties. As it happens, standard speech is everywhere the most constrained of all varieties, that is, the one that imposes the tightest restrictions upon variation. When we look at other social varieties, we invariably find loosening of constraints. Working-class varieties in all communities, for instance, tolerate (CC) simplification in more formal stylistic registers, including technical talk and interviews. Traditional vernaculars go further, relaxing linguistic constraints as well as stylistic ones.

Studies of (C̲C) variation in traditional vernaculars, including, for instance, unrelated varieties such as inner-city African American (Guy 1980) and rural northern England (Chambers 1982), though separated geographically by the Atlantic Ocean and separated socially by nationality, urbanity, and ethnicity, draw the line very differently from the standard varieties but virtually identically to one another. Socially, both the African Americans and the rural Englishmen are self-contained subcultures relatively isolated from standard or middle-class norms. Stylistically, they permit (C̲C) simplification in formal speech as well as casual. Linguistically, looking at Table 24.1 again, they do not draw the line anywhere, that is, (C̲C) simplification can take place in any of the cells, whether before a consonant (*lan' fall*) or a vowel (*lan'owner*) or finally (*lan'*). Variation is not, however, free (or random). There is a quantitative regularity so that (C̲C) simplification occurs more frequently before a consonant (the first row of Table 24.1) than elsewhere. That is, the linguistic context that is the only one permitted in middle-class varieties is also the one that is most frequent in vernacular varieties.

Phonotactically, where (C̲C) simplification in standard accents effectively simplifies triconsonantal clusters, in vernaculars it simplifies biconsonantal as well as triconsonantal clusters. The *reductio ad absurdum* of variable processes like (C̲C) simplification would be self-annihilation. If in some variety consonant clusters were invariably reduced, they would cease to exist in the dialect structure. That *reductio* state may have been attained in certain African American varieties in which nouns like *desk* and *test*, pronounced *des'* and *tes'*, occur in those varieties with plural forms *desses* and *tesses*, as if the lexical singular were *dess* and *tess*. For those varieties, cluster simplification may have become lexicalized, simplifying the syllable structure by eliminating final clusters and thus eliminating the variable process (C̲C).

3 Social Correlates

Recognizing the linguistic variable as a structural unit led to several innovations in linguistic analysis. The most obvious departure from venerable traditions is the incorporation of social attributes as independent variables. Linguistic conditioning had always figured in phonological distributions, including determination of allophones and allomorphs, and it remains an important independent variable in sociolinguistic analysis, as in the probabilistic conditioning of (C̲C) simplification in the presence of a following consonant compared to a following vowel. Social attributes had formerly been ignored as systematic conditioning factors. However, they demonstrably exercise probabilistic influences as conditioning factors on variation, and must therefore be admitted into the formal analysis. The core social attributes affecting language use are social class, social networks, sex and gender, ethnicity, and age. Sociolinguists have discovered, as Edward Sapir so eloquently predicted more than three decades before sociolinguistics came into being, that "behind the apparent lawlessness of social phenomena there is a regularity of configuration and tendency which is just as real as the regularity of physical processes in a mechanical world" (1929: 211).

3.1 Social classes

Social classes are abstract stratifying categories primarily determined by the intersection of education, occupation, and income. People with similar attainments in these matters tend to socialize with one another, live near one another, and come to share attitudes, recreations, and aspirations. In industrialized societies, the major class divisions are middle class and working class. Upper-class constituencies, the stratum of people with inherited wealth and/or status, have been diminishing in numbers and influence for over a century and are seldom considered. They always were minorities. In Europe and parts of Asia, social structures inherited partly from feudal traditions resulted in pyramidal hierarchies so that upper-class members, the apex

of the pyramid, exerted a disproportionate influence socially, economically, and linguistically. Those traditions were still visible in much of Europe until the middle of the twentieth century. In Great Britain, for instance, the prestige accent was Received Pronunciation, popularly known as the Queen's (or King's) English or the public school (or Oxbridge) accent, after its archetypal members, who never amounted to more than 5 percent of the population. From 1950 onward, the global spread of New World social and occupational mobility has eroded upper-class prestige, with predictable sociolinguistic consequences. In England and elsewhere in Europe, standard dialects are increasingly middle-class dialects, as they always have been in New World nations, and upper-class dialects are themselves importing middle-class features to the point where they are nearly indistinguishable.

The main divisions, middle class (MC) and working class (WC), grossly characterized as white collar and blue collar, are made up of a finely graded continuum in complex urban communities with lower, middle, and upper subdivisions from lower working class to upper middle class. Together, the continuity of MC and WC dialects is recognized by calling them "mainstream," a useful term put forward by the dialectologist Wells (1982: 4) to distinguish them from "traditional" dialects, the radically different, often noncontinuous varieties found in ethnic or geographic enclaves (including, for instance, inner-city African American and rural northern British vernaculars). The middle-middle-class variety is typically the standard; by definition, its speakers are not identifiable regionally by their accents. Movement down the social hierarchy is marked linguistically by increasing use of nonstandard variants and regional variants. These differences in social dialects are typically quantitative rather than qualitative. Thus, as we have seen, standard MC varieties have, say, (C<u>C</u>) simplification under well-defined linguistic and stylistic conditions, and those conditions, though they remain well defined, are relaxed by degrees in the speech of each successive subclass. This gradation of quantitative increments is typical of phonological variables like (C<u>C</u>). Grammatical variables can also be graded down the social hierarchy, but they are often qualitative in the sense that they occur in WC speech but not in MC speech, or at least not in the MMC standard. Well-known grammatical variables that almost never occur in MMC speech but occur with increasing frequency down the WC hierarchy are multiple negatives (*I didn't do nothing*), conjugation regularizations (*I done it*, *I seen it*, *He come late*), and pronoun plural *youse*. Grammatical variables like these thus constitute markers of class dialects.

One of the discoveries that arises from considering social class as an independent variable is the predictable structural relationship between standard and nonstandard varieties. The standard variety always is more rigidly constrained (as we saw for (C<u>C</u>) simplification). Generally, the standard variety admits fewer phonological coalescences and requires more elaborated grammatical agreement and concord. These differences have no *linguistic* consequences: that is, there is no semantic or pragmatic difference between, for instance, standard *I didn't do anything* and nonstandard *I didn't do nothing*. They do, however, have *social* consequences, with the former evaluated as normal, educated, and mannerly but the latter as rough, uneducated, and impolite.

In other words, the distinction between standard and nonstandard varieties universally carries social, not linguistic, significance. Predictably its origins are social. Kroch was the first to make this observation in the sociolinguistic context. "Dominant social groups tend to mark themselves off symbolically as distinct from the groups they dominate and to interpret their symbols of distinctiveness as evidence of superior moral and intellectual qualities," he said (Kroch 1978: 18). "This tendency shows itself not only in speech style but also in such other areas of social symbolism as dress, body carriage, and food." This insight is now integrated into the theory known as "vernacular roots" (Chambers 2009: 258–66), the implications that follow from the discovery of apparently universal structural relationships in the standard-to-vernacular hierarchy and the recurrence of primitive variable processes in the vernacular forms of all languages.

3.2 Social networks

Social networks are microlevel social clusters such as neighborhoods, parishes, institutions, and clubs. They resemble social classes in the sense that they are made up of people with similar social attributes, but they are more narrowly construed and more intimate. In all communities, people vary in the extent of their involvement in local networks. They range from highly integrated at one extreme to loosely engaged at the other, differences characterized in everyday language by such terms as clannish or cliquey at one end to reclusive or alienated at the other. Most people are neither cliquey nor recluses but somewhere in between. Individual participation in networks is measurable in terms of density and multiplexity, where density is measured by the frequency of contact a person has with individuals in an identifiable cohort, and multiplexity is measured by the number of bonds shared with those individuals (siblings, neighbors, workmates, recreational partners, and so on). Middle-class adults in urban societies typically maintain loose network ties. Often, middle-class adults leave the homestead to settle among relatively transient apartment-dwellers or suburbanites, where they commute to work, enroll children in schools beyond the home community, and visit relatives or old acquaintances infrequently. Other individuals spend their days and nights at closer quarters, and for these people the density of the network has linguistic consequences (Milroy and Llamas 2013; Chambers 2009: 74–92).

Socially, networks are norm-enforcement mechanisms; this notion is perhaps more explicit in the term "community of practice," commonly used nowadays in network studies, defined as a group that comes together around a particular enterprise (Meyerhoff and Strycharz 2013; Eckert 2000). Norm enforcement can influence attitudes, recreations, deportment, dress, hairstyles, and other aspects as well as speech. Adolescence is typically a phase of intense network bonding, with considerable social and linguistic consequences, as we shall see when we consider age as an independent variable.

Social networks have proven indispensable in the study of insular groups of all kinds, including African American gangs in Harlem, Alpine villagers in Austria, working-class enclaves in Belfast, and aristocrats in Main-Line Philadelphia. Even in highly circumscribed communities like these, an individual's participation in the network can vary from dense to relatively loose. It is possible to place individuals on a continuum using a checklist that reflects local social practice. In Belfast, for instance, Milroy devised a Network Strength Scale by allotting one point for each condition that a subject satisfied: (1) membership in a local club, such as a bingo hall, football team, or card-playing group, (2) kinship ties with two or more households in the neighborhood; (3) work in same place as two or more neighbors; (4) work alongside two or more neighbors of the same sex; and (5) association with workmates outside the workplace. Individual were ranked from 0 to 5 on the Network Strength Scale, with the higher scores representing greater density and multiplexity.

The sociolinguistic pattern is clear in the Belfast enclaves and all other tight communities. The higher the Network Strength Scale, the greater the use of local vernacular variants, and vice versa. In the insular communities for which network studies are most revealing, there is a high degree of homogeneity among citizens in terms of the macrolevel indicators like occupation, income, and education. It follows that sociolinguistic analysis in terms of, say, social class should reveal considerable homogeneity. Yet language variation persists as hardily in these communities as in more fluid ones. One of the most important sources is the gradation of linguistic conformity to local dialect norms correlated closely with the individual's integration into the community.

3.3 Sex and gender

Sex and gender are related but separable attributes. The case for distinguishing between sex and gender was first made by feminist linguists in the 1970s and has proven to be highly productive. Miller and Swift (1976: 51) said, "At the risk of oversimplification, sex . . . is a biological

given; gender is a social acquisition." Sex includes the complex of physiological, neurological, and cognitive differences that distinguish females and males. Gender is the complex of cultural and social roles deemed appropriate for females and males; it is learned or inculcated starting in childhood, and it can differ from one culture to another.

Both sex differences and gender differences can affect speech. A sex difference that obviously affects language is the lengthening of the glottis in males at puberty, which results in men having, on average, lower-pitched speaking voices than women. A gender difference that affects language in all English-speaking cultures (and many others) is that women use fewer swear words (*shit*, *cunt*, etc.), on average, than men of the same social class and age. The sex difference is determined by biology, and the gender difference by social convention.

Sex differences are visible from conception but become more visible from puberty onward. Gender differences are largely invisible. Most sociolinguistic studies make correlations with sex groups, that is, dividing subjects into females and males, without regard for gender roles that might exist. That expedient hardly affects the sociolinguistic analysis of middle-class subjects in advanced economies because conventional roles for women and for men at this social stratum are almost indistinguishable.

Gender roles in some societies are sharply differentiated and enforced, and in those societies linguistic differences between women and men are equally sharp. At the most extreme pole, in a few societies women and men are said to speak different languages. That is always an overgeneralization based on anthropological (not linguistic) analysis, but it highlights the extent to which their speech differs. The most famous case is Carib, spoken in the Lesser Antilles. According to a seventeenth-century adventurer, "The men have a great many expressions peculiar to them, which the women understand but never pronounce themselves, [and] the women have words and phrases which the men never use, or they would be laughed to scorn" (quoted in Trudgill 2000: 65). The differences mainly involve lexical exclusions but they also affect self-reference forms and certain phrase types, yielding grammatical differences, and in some languages extend to phonology. The women and men are actually bidialectal in terms of perception but social conventions forbid one sex group from using linguistic forms deemed appropriate to the other. The differences are great enough and the conventions are so rigidly enforced that outsiders understandably mistake them as being bilingual rather than bidialectal. Extreme cases like these are known as "sex-exclusive variation," and they have been found only among technologically undeveloped peoples that enforce rigid differences between female and male roles.

In mainstream societies, male and female differences are sex-preferential rather than sex-exclusive. The use of certain variants by the sexes is a matter of degree, not kind. Where sex-exclusive variation in primitive societies is inviolable, in mainstream societies there is said to be no prohibition on members of the opposite sex choosing variants characteristically used by the other sex. In practice, however, there is definite, predictable stereotyping, often accompanied by censure, at least in the most extreme cases. For instance, a woman would be marked as "rough" if she used a masculine curse-word like "cocksucker," and a man would be marked as effeminate for using a feminine adjective like "darling." Men and women occasionally do make these choices, of course, presumably to evoke the stereotype, often humorously but not always.

In mainstream societies, differences in sex-preferential variation are often class-based, especially between middle-class and other strata. Working-class communities in Belfast, studied by James and Lesley Milroy, and inner-city communities in Detroit, studied by Walt Wolfram (reviewed and discussed in Chambers 2009: 125–36), reveal significant female–male differences for every linguistic variable, both phonological and grammatical. The direction of variability is the same in both communities: Women use fewer stigmatized and nonstandard variants and command a wider range of speech styles than do the men who are their brothers, husbands, and neighbors. The difference is explained by *gender-based variability*, because it follows from well-defined differences in gender roles. In these communities, the men are much more circumscribed in their movements, usually working locally and socializing within the neighborhood. Their

insularity comes about partly because of sectarian or racial pressures, which are thought to make it dangerous for men to venture outside their home territory. Women can venture further with impunity, and often work, shop, and develop friendships beyond their neighborhoods. It follows from these differences that women acquire greater sociolinguistic range as a result of their occupational and geographic mobility.

With such a ready and reasonable explanation in terms of gender-based variability, it is surprising to discover that female–male differences also occur in middle-class communities, where gender-role differences are vague at best and nonexistent in any meaningful sense in contemporary technologically advanced societies. Equally surprising, the direction of variability is exactly the same: Middle-class women use fewer stigmatized and nonstandard variants and command a wider range of speech styles than do middle-class men, that is, the men who are their brothers, husbands, and neighbors. The differences are more gradual than in the lower classes, and the variables are inevitably fewer because some grammatical variables rarely or never occur in middle-class speech, but they nevertheless show up with statistically significant frequency in careful studies all over the world. These differences are clearly not gender-based because of the mild or nonexistent gender differences. Instead, they are hypothesized as *sex-based variability*, related to the verbal advantage that women show in standardized tests of fluency, sentence complexity, analogy, listening comprehension of both written and spoken material, vocabulary, and spelling (Chambers 2009: 141–2, 145–51). That verbal advantage presumably follows from sex-based neural differentiation, what neurologists call "sexual dimorphism in brain asymmetry." That dimorphism also accounts for greater female immunity from stuttering, aphasia, and dyslexia as well as numerous nonlinguistic tendencies such as right-handedness. The differences reflect statistical tendencies – there are, of course, left-handed women who stutter, just significantly fewer of them than men. The sociolinguistic advantage of women over men, which is also a tendency and a statistically significant one, appears to be another manifestation of these inherent differences.

Sex-preferential differences follow nonarbitrary, largely predictable trends in every community where they have been studied systematically. The strength of those trends apparently follows from the fact that they are based on both gender and sex differences. However, it goes deeper. Gender-based variability appears to be a concomitant of sex-based variability. In societies where gender-based variability is a factor, that is, in societies in which the roles of women and men are sharply defined, it is invariably women, not men, who are assigned the linguistic "chores" – the women, not the men, are the ones who make appeals to the landlord, represent their children's interests with school officials, negotiate credit with bank officers, and otherwise speak on behalf of the family in public. This asymmetry is surely no accident. Presumably women take on the linguistic chores in communities where those chores are gender-specific because they are predisposed to do them by their innate sociolinguistic advantage.

3.4 Ethnicity

Ethnicity is actually subsumed under social class and network analysis in urban societies. Where ethnic enclaves exist, as in inner-city African American communities, immigrant-based neighborhoods called Chinatown, Little Italy, and the like, or sectarian parishes for Hasidic Jews or Ukrainian Catholics, distinctive linguistic features often come into being as a function of social segregation on the one hand and identity construction on the other. Ethnicity is never a determining factor for dialects by itself, but ethnicity sometimes determines the social circles a person interacts with. When it does, linguistic variants associated with the ethnic group are acquired exactly as are variants associated with class and network. As in any other network, people who live in ethnic enclaves differ with respect to their integration into it, and those who are loosely tied do not acquire the vernacular norms of the enclave as strongly as those who are tightly integrated.

Ethnic enclaves, like all networks, are norm-enforcement mechanisms. People who work or study outside the enclave but live inside it necessarily develop facile stylistic accommodation in their speech, because peer pressure within the enclave evaluates their "outside" accent as "putting on airs." In ethnic enclaves, individuals vary considerably in the range of their sociolinguistic behavior, as in fact do all communities. This fact became obvious early in the sociolinguistic enterprise and quickly dispelled the naive idea that speech communities have to be linguistically homogeneous. We now know that they are linguistically diverse but are in fact united in their linguistic attitudes. One of the abiding results of William Labov's seminal study of the linguistic stratification of New York City was his discovery that "the speech community is not defined by any marked agreement in the use of language elements, so much as by participation in a set of shared norms" (Labov 1972: 120–1). Members of the community subconsciously agree on the range of tolerable variation and its social significance.

Ethnicity attracted disproportionate attention as an independent variable during the formative years of sociolinguistics. Numerous concepts had their inception in Labov's work in the African American community in New York and studies it inspired in inner-city Detroit, Washington, Philadelphia, and elsewhere. At the beginning of the twenty-first century, researchers estimated that there were at least five times as many sociolinguistic studies devoted to African American groups as to all other ethnic groups. This bias proved to be a salutary starting point because the linguistic distance between African American vernacular and other English varieties afforded an extreme and relatively well-defined perspective on theoretical concepts. It also instilled a bias for studies of vernaculars in general, the natural, unselfconscious, uncodified talk that gives most direct access to the language faculty and provides a baseline for understanding constraints on standard and literary dialects.

3.5 Age

Age has probably proven the most productive independent variable because it provides a window on language change. Gauchat, decades before sociolinguistics became a discipline, said, "Variation in pronunciation among members of a single speech community has not been studied systematically, despite its potential contribution to our understanding of language change" (2008 [1905]: 227]). Some six decades later, Weinreich, Labov, and Herzog declared,

> A model of language which accommodates the facts of variable usage and its social and stylistic determinants not only leads to more adequate descriptions of linguistic competence, but also naturally yields a theory of language change that bypasses the fruitless paradoxes with which historical linguistics has been struggling for half a century. (1968: 99)

Sociolinguistic theory provides a dynamic view in which change is apprehended in progress, so that leaders and laggards can be identified and both the course of its diffusion and its rate can be delineated.

The crucial theoretical innovation is the apparent-time hypothesis. Saussurean or structuralist theories of language change maintained that change could only be apprehended in real time, that is, by comparing the language in two states (or more) separated by a time interval. Sociolinguistic theories, by contrast, view change in apparent time, by comparing speech of people of different ages in the community and inferring from observed differences that linguistic change is in progress. Direction of change is inferable under normal circumstances by observing that younger speakers show a preference for a particular variant that is not present or (more often) less frequent in the speech of their elders. These observations about diachronic change can thus be made synchronically, that is, in apparent time.

Inferences of change based on apparent-time evidence depend crucially upon the axiom that individuals generally maintain throughout their lives the speech norms they acquire in their

formative years, from 8 to 18. The axiom holds almost inviolably except for a few well-understood exceptions such as age-grading, which involves the systematic replacement of "childish" variants at maturity (as when childhood words like *Mommy* and *pee-pee* are replaced by adult words); in those cases the speech of the younger group is not innovative but a kind of maturational phase. There are potentially other disruptions –interruption of a change in progress by a competing change, stasis due to stereotyping of the incoming variant, and the like – but they are rare. They do happen, and for that reason inferences based on apparent time must always be tested. Nevertheless, apparent time has provided robust evidence for unraveling the multivariate complexities of language change.

Cumulative evidence shows that prototypical language changes are accelerated by adolescents and led by women. Adolescence is a period of tight-knit network affiliations, even in the middle class in which most adults maintain relatively loose networks. Adolescence is a transitional phase in which teens increasingly distance themselves from adults, especially parents, as preparation for their own impending independence; at the same time, apparently as compensation, bonds with peers become denser. Adolescents are particularly inclined to adopt innovations in fashion, including dress, diet, social causes, hairstyles, and decorative fads of many kinds. Less superficially, they also adopt innovations in their speech style, most obvious in teenage slang but also, subconsciously, in accent features, where adolescents can be observed increasing the frequency of certain variants above the communal norm. Peer pressure, as in any tight network, tends to spiral the new norms upward. After adolescence, retrenchment is normal; teen slang may be avoided and tattoos may be surgically removed, for instance, but more subtle innovations such as the new target for a certain vowel, because beneath consciousness, are likely to persist.

Women lead changes that are standardizing. Men, not surprisingly, lead changes that involve the rise of colloquial variants. For instance, the spread of the affirmative "okay" in southern England, once shunned as an Americanism, is being led by English men. Regardless of who leads the changes, they are normally incremental. Viewed across a long interval, for instance in the 60-year interval between 20-year-olds and 80-year-olds, changes may create what appears to be a generation gap, but the "gap" is an artifact of viewing change as a static phenomenon. The speech of the people between the 20-year-olds and 80-year-olds normally represents a gradual progression. Speakers in their 70s use variants much like the 80-year-olds but with slightly more of the incoming variant, and so on in steps down the age continuum. Intermediate age groups are indeed intermediaries. Because of this, linguistic changes are seldom disruptive. The rate of change is governed by a self-regulating principle of generational compatibility. In every society, personal relations are more intimate with people similar in age – an individual's classmates, workmates, team-mates, and partners are usually people within about five years, one way or the other. The fact that they share vernacular norms with one another is the result of their social proximity, and also its cause.

Incremental changes are gradual but they are not necessarily evenly spaced along the age continuum. Normal linguistic changes follow the S-curve model, just like innovation diffusions in technology, fashions, and other matters. The S-curve is a metaphor based on the graphic representation of changes, which trace a pattern that is relatively flat at both start and finish, and linked by a rapid rise in the middle. (The metaphoric S must be made oblique.) In the progress of changes from start to finish, they start slowly for about the first 25 percent until the innovative form attains a critical mass in frequency, and at that point it accelerates rapidly through the middle 50 percent; there is then a tailing off as the change nears completion, so that the curve flattens for the final 25 percent, normally with some lingering residue of older forms.

Studies of changes in progress might apprehend a particular change at any phase of the S-curve. The Dialect Topography of Canada survey fortuitously caught several changes from inception to completion in its 70-plus years of apparent-time from the 1920s to 1990s (Chambers 2013b: 360–70). Two such changes are the merger of /hw/ with /w/ in words like *why*, *whine*, and

whale and the replacement of *sneaked* by *snuck*, the past tense of *sneak*. As expected, the changes go through a flat, relatively slow beginning, and then leap forward in the accelerated middle phase until they tail off as they near completion. In the accelerated phase, progress is greatest in the decade in which the incoming variant became majority usage. In this leap-forward phase, increments can rise as high as 20 percent, but they are otherwise more or less even in the accelerated phase, in keeping with generational compatibility. Dating the phases of the change – its relative chronology, in traditional terms – can be accomplished with considerable precision simply by calculating birth dates from subjects' ages at the time of the study: People in their 40s in a 1992 survey were born in between 1943 and 1952. Of seven major changes in the Canadian survey, the leap-ahead phase for all of them happened in the 1940s or 1950s. These were culturally significant years in Canada marked by the diminution of British influence with the dissolution of the Empire and by multiethnic post-War immigrations that diluted the Anglo-Celtic hegemony of the country.

The Canadian case study, like many other studies in which sociolinguistic dimensions have been fastidiously integrated, establishes beyond a doubt the primacy of social motivation in language change. The seven changes that move toward completion almost simultaneously are linguistically diverse, involving syntax, morphology, phonology, pronunciation, and lexicon. They are also diverse in their provenience, with some of them ancient variations (such as /hw/ merger, incipient in the very earliest records of the English language) and some of them recent (such as *snuck* for *sneaked*, first attested in Louisiana in 1887 but unknown in Canada until the mid-twentieth century). Linguistic motivation for the changes ranges from functional (/hw/ is the only preaspirate in the English consonant inventory since /hr/ and /hl/ were eliminated in the Middle Ages) to bizarre (*snuck* replaces a weak form with a strong one, with no analogical precursor). Notwithstanding the linguistic diversity in terms of structure, provenience and motive, these changes progress at the same historical moment, lock-step. They do so in response to sociocultural changes that transformed the country starting in the middle of the twentieth century.

Sociolinguistic perspectives on linguistic change have brought unprecedented clarity to our understanding of its actuation, implementation and motivation. In retrospect, it has demystified some venerable puzzles and is capable of demystifying more. Based on the premise that the forces underlying linguistic change today are the same since time began, a linguistic application of the time-honored uniformitarian principle that underlies all Western science (Labov 1972: 275), old puzzles simply vanish. Claims that Grimm's Law had no exceptions, for instance, or that the Great English Vowel Shift was completed virtually overnight are now recognized as artifacts of narrow methodologies. With hindsight, it is clear that evidence for variation and generational compatibility was formerly shunted aside as "free" variation in diachronic studies or as scribal error in philological studies. Now they are recognized as vernacular variants, and we see them as the legitimate witness to variable usage by a broader social spectrum.

4 Theory and the Accidents of History

The social sciences are relatively recent developments in intellectual history. The idea that human behavior could be studied with precision comparable to the physical sciences came late. Anthropology, economics, sociology and the other social sciences had their beginnings in the second half of the nineteenth century. By then, natural sciences like physics, chemistry, and geology had more than a millennium head start. Sociolinguistics, the social science of language studies, was a belated addition even among its late-blooming sister disciplines. Its effective beginning as a discipline with common goals and international practitioners dates from the dissemination of William Labov's ideas in the 1960s. Nonsocial language studies such as grammar, philology, and lexicology, by contrast, are as old as physics and chemistry.

Modern linguistics itself had undergone sophisticated evolution in the twentieth century before sociolinguistics came into being, starting with Saussure's posthumously published lectures in 1916 through American and European schools of structuralism to Chomsky's mentalist model. Before Labov, all linguistic theories shared the idea that linguistic units were invariant, discrete and qualitative. Sociolinguistics represented a revolutionary insurgency in theoretical terms. Historically it was not received as revolutionary. Labov recalled his surprise at the positive reception to his first presentation to the Linguistic Society of America in 1963. "I had imagined a long and bitter struggle for my ideas, where I would push the social conditioning of language against hopeless odds, and finally win belated recognition as my hair was turning gray," he recalled (1997). "But my romantic imagination was cut short. They ate it up."

Sociolinguistics was timely, perhaps overdue. One of the consequences of its easy reception was rapid progress both in the development of seminal ideas and their diffusion around the world.

Not surprisingly, some areas remain conspicuously underdeveloped. Vernacular studies are almost exclusively based on English varieties, with a conspicuous lack of studies in, say, working-class French or German or Japanese or any other language. One possible explanation is that scholars in older, more hidebound academic systems are drawn from an intellectual caste that mixes uneasily with lower strata; if so, this hiatus will be shortlived as occupational mobility continues to spread globally. Another underdeveloped area, probably not unrelated, is comparative sociolinguistics, in which results from different sociocultural settings are examined together to yield explicit, replicable hypotheses about sociolinguistic patterns beyond single linguistic communities. Truly significant insights into sociolinguistic competence as a module of the innate language faculty will probably depend upon comparative identifications across language boundaries, and on vernacular processes shared by English and German and Japanese and other languages. That gap in our knowledge may take longer to fill, and in fact is not foreseeable.

Another consequence of the rapid spread of sociolinguistics as an international discipline is increasing diffuseness of its disciplinary mandate. Its success is rooted in its social science base but it has attracted practitioners with qualitative inclinations, more inclined toward cultural anthropology than linguistics, and others in peripheral areas such as pragmatics where confidence in any scientific sense is hard to come by. Their findings may be salutary in so far as they are complementary to the sociolinguistic enterprise, but their relevance is often unclear. The core of sociolinguistics, if it is to retain its stature in the discipline, must be the social evaluation of linguistic variants, the area most susceptible to scientific methods such as hypothesis formulation, logical inference, and statistical testing.

In the long history of systematic language studies from Pāṇini and classical Greece to the midpoint of the last century, linguistics was nonsocial and occasionally antisocial. Sociolinguistics participates in analyses of grammar, phonology, and phonetics along with other branches of its discipline. Beyond those shared analytic concerns, it has developed endemic theories hitherto unexplored. Principal among them is the merging of synchrony and diachrony, obliterating the old Saussurean artifice, in recognition of the obvious truth that (as Varro said more than two thousand years ago) "the vernacular is always in motion." Its motion is palpable to the members of any community in the variation among social groups, especially people differentiated by age, class, networks, and sex, the main stratifying attributes in societies, and, it follows, the main correlates of linguistic variation and change. It is also recognized by individual participation in social contexts, the continuum called style, because each participant embodies to greater or lesser extent the range of the communal variation. Like grammatical competence, our sociolinguistic competence is neither taught nor consciously learned. It is simply acquired by virtue of our human nature, and deeply embedded in it. Sociolinguistics seeks to discover its properties, describe its mechanisms, and ultimately understand its place in the language faculty.

SUGGESTIONS FOR FURTHER READING

Chambers, J. K. and Schilling, N. (eds) (2013). *The Handbook of Language Variation and Change*, 2nd edn. Oxford: Wiley Blackwell.

Labov, W. (2001). *Principles of Linguistic Change, Vol. 2: Social Factors*. Oxford: Blackwell.

Tagliamonte, S. A. (2006). *Analyzing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

REFERENCES

Basilius, H. (1952). Neo-Humboldtian ethno-linguistics. *Word*, 8, 95–105.

Chambers, J. K. (1982). Geolinguistics of a variable rule. *Discussion Papers in Geolinguistics*, 5, 1–17.

Chambers, J. K. (2008). Louis Gauchat, proto-variationist. *Historiographica Linguistica*, 35, 215–25.

Chambers, J. K. (2009). *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*, 3rd edn. Oxford: Wiley Blackwell.

Chambers, J. K. (2013a) Studying language variation: an informal epistemology. In J. K. Chambers and N. Schilling (eds), *The Handbook of Language Variation and Change*, 2nd edn (pp. 1–15). Oxford: Wiley Blackwell.

Chambers, J. K. (2013b) Patterns of variation including change. In J. K. Chambers and N. Schilling (eds), *The Handbook of Language Variation and Change*, 2nd edn (pp. 297–323). Oxford: Wiley Blackwell.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1980) *Rules and Representations*. New York: Columbia University Press.

Eckert, P. (2000). *Linguistic Variation as Social Practice*. Oxford: Blackwell.

Gauchat, L. (2008 [1905]). Phonetic unity in the dialect of a single village, translated by S. Cummins. *Historiographica Linguistica*, 35, 227–75.

Guy, G. (1980). Variation in the group and individual: the case of final stop deletion. In W. Labov (ed.), *Locating Language in Time and Space* (pp. 1–36). New York: Academic Press.

Hjelmslev, L. (1961). *Prolegomena to a Theory of Language*, translated by F. J. Witfield. Madison, WI: University of Wisconsin Press.

James, W. (1911). *Some Problems of Philosophy*. London: Longmans, Green.

Joos, M. (1950) Description of language design. *Journal of the Acoustical Society of America*, 22, 701–8. (Reprinted in 1957 in *Readings in Linguistics I: The Development of Descriptive Linguistics in America 1925–56* (pp. 349–56). Chicago, IL: University of Chicago Press.)

Kroch, A. (1978). Toward a theory of social dialect variation. *Language in Society*, 7, 17–36.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.

Labov, W. (1997) How I got into linguistics, and what I got out of it." Undergraduate address 1987 revised. Available online at http://www.ling.penn.edu/~labov/Papers.

Meyerhoff, M. and Strycharz, A. (2013) Communities of practice. In J. K. Chambers and N. Schilling (eds), *The Handbook of Language Variation and Change*, 2nd edn (pp. 428–47). Oxford: Wiley Blackwell.

Miller, C. and Swift, K. (1976). *Words and Women*. Garden City, NY: Anchor Press, Doubleday.

Milroy, L. and Llamas, C. (2013) Social networks. In J. K. Chambers and N. Schilling (eds), *The Handbook of Language Variation and Change*, 2nd edn (pp. 409–27). Oxford: Wiley Blackwell.

Sapir, E. (1921). *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace.

Sapir, E. (1929). The status of linguistics as a science. *Language*, 5, 207–14.

Saussure, F. de (1916). *Course in General Linguistics*, edited by C. Bally and translated by A. Sechehaye and W. Baskin. New York: McGraw-Hill.

Trudgill, P. (2000). *Sociolinguistics: An Introduction to Language and Society*, 4th edn. London: Penguin.

Weinreich, U., Labov, W., and Herzog, M. I. (1968) Empirical foundations for a theory of language change. In W. P. Lehmann and Y. Malkiel (eds), *Directions for Historical Linguistics: A Symposium* (pp. 95–188). Austin, TX: University of Texas Press.

Wells, J. C. (1982) *Accents of English 1: An Introduction*. Cambridge: Cambridge University Press.

25 Language Variation

Sociolinguistic Variationist Analysis

KIRK HAZEN

1 Introduction

The study of language variation in linguistics is a hybrid enterprise. In some respects, it is a branch of sociolinguists, examining the ebb and flow of language in society. In other respects, it is more closely aligned with linguistics, developing explanations and models for complex linguistic systems. From this diverse work, the study of variation is often connected to the communities themselves, and scholars have worked to rectify false perceptions of vernacular dialect varieties. Not all scholars who pursue these different paths balance their time between them evenly, but the study of language variation is important to all of them.

As expected given its hybrid nature, language variation scholarship springs from a complex lineage (see Hazen 2007a for a fuller history). Dialectology, historical linguistics, generative grammar, and applied linguistics have all played a role in how language variation is studied today. From the most global perspective, many linguists can be said to pursue the study of language variation. As human language boasts over 6,900 living varieties, most of the pressing questions in linguistics revolve around the variable qualities of human languages, and extensive bilingualism and language contact increase that variability (Sankoff 2013). The study of language variation works in harmony with this goal by pursuing synchronic and diachronic explanations of how language variation works in society and in the mind (Fasold 1990; Labov 1994).

This chapter introduces some of the range of variationist studies, including examples from recent research, and describes interactions with other linguistic fields. The chapter also highlights some of the community outreach that variationists perform as a result of their scholarship. It is organized around the following themes: the range of language variation, divisions of language variation, the locus of language variation, constraints on language variation, quantitative analysis, and then the linguistic systems where language variation is frequently studied, including the lexicon, morphology, syntax, phonology, and phonetics. Throughout the chapter, qualities of variationist sociolinguistics are exemplified with modern studies. As with most areas of linguistics in the twenty-first century, the scholarship is so voluminous that many other high-quality studies could have been used as examples in this chapter.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

2 The Range of Language Variation

Linguists generally divide language variation between two overarching areas: *diachronic variation* and *synchronic variation*. At its most basic, diachronic variation covers the differences found in one language variety at two different points in time (Hock and Joseph 1996). In the nineteenth and the first half of the twentieth century, scholars examined diachronic variation over the span of centuries. For example, William Dwight Whitney (1867) studied branches of Indo-European, such as Sanskrit, to trace their developments. Such diachronic studies were the predominant form of language variation analysis until a shift in linguistic priorities, fostered by Ferdinand de Saussure (1916) and propagated in the United States by Leonard Bloomfield (1933), turned linguists' attention to synchronic study. Scholars who take up synchronic language variation examine the qualities of one or more varieties at a single point in time. The study of language variation in the early twentieth century greatly expanded linguistic knowledge of the vast, yet systematically constrained, range of human language. The extent of diversity led some to scholarly despair, as their attempts to describe categorical rule-based systems failed consistently in the face of variable data. Edward Sapir (1921: 38), renowned scholar of American Indian languages such as Navajo and Yana, remarked: "Unfortunately or luckily, no language is tyrannically consistent. All grammars leak." Many linguists in the twentieth century attempted to fix the leaks by ignoring language data through processes of logical abstraction. In other words, they ignored the leaks and claimed that only the dry places were worthy of study. The modern study of language variation was developed by focusing on the variation which made the grammars leak.

Although others examined language variation in the 1960s, Labov's (1963) study of Martha's Vineyard was the first to directly connect diachronic variation with synchronic variation. In contrast to previous scholars, Labov argued that language change was observable from synchronic variation. This innovation began a minor revolution in linguistics wherein synchronic language variation became the direct focus of study in order to explain quandaries of language change, social change, social groupings, and dialectology (e.g., Labov 2001, 2006). The study of language variation and change became its own subfield because of Labov's efforts, and he has continued as a leader of the field for half a century (Hazen 2011a).

3 Dividing the Landscape of Language Variation

The efforts of language variationists cover a wide range of linguistic disciplines (Wolfram 2006). Although variationists do not generally label themselves by the categories below, these categories illustrate the precipices from which we can make observations of the modern work on language variation in linguistics. The full range of sociolinguistics is even broader, and several of its goals, including the focus on social and identity research, are not explored here (cf. Holmes and Hazen 2014; Wodak, Johnstone, and Kerswill 2011).

Scholars do not necessarily agree about how to classify all linguistic variables, and the criteria for such classifications are part of the developing nature of the field of language variation and change. Two examples illustrate the arguments made for different linguistic boundaries. In the first example, US Southern varieties of English, including the related African American Vernacular English (AAVE), can replace *there* in a phrase like *there's a lot of them* with *it* as in *it's a lot of them*. Some scholars have referred to this as syntactic variation, since the existential subject is being replaced; however, if a variation study finds no syntactic factors affecting the variation, it may also be lexical variation. The second example takes up one of the staples of variationist analysis: (ING).[1] First investigated by Fischer (1958), this sociolinguistic variable continues to yield information about many aspects of language variation (e.g., Campbell-Kibler 2007, 2008, 2011; Hazen 2008). The variation resides between two forms, an alveolar nasal form (e.g., *walki*[n])

and a velar nasal form (e.g., *walki*[ŋ]). Despite numerous studies, early researchers could not form a consensus on what kind of variable it is. For most of its history, it was seen as a phonological variable (sometimes called *velar fronting*), with the assumption that some factor was triggering a change in the place of articulation from the velar to the alveolar nasal. However, few studies found any phonological influence (cf. Tagliamonte 2004). Eventually, Houston (1985) and Labov (1989) made successful arguments that (ING) variation results from morphological and social pressures, with the gerund taking the velar nasal more often (e.g., *walki*[ŋ] *is fun*) and the verbal progressive taking the alveolar more often (e.g., *we were walki*[n]). This complexity of factors fits well with Wolfram's assessment of the linguistic variable. He writes:

> Though sensitive to traditional structural categories and relationships in language, the linguistic variable is not necessarily beholden to them, and there are clearly cases where the imposition of a traditional linguistic boundary may actually detract from rather than enhance the understanding of language variation in its social context (Wolfram 1993). Such considerations may be one of the reasons why it might be argued that the linguistic variable is a uniquely sociolinguistic construct that conveniently – if not always comfortably – covers the locus of co-variation between linguistic variants and other variables that include social, historical, psychological, and linguistic factors (He 2006: 336).

Hence, the (socio)linguistic variable is different from theoretical tools found in traditional linguistic disciplines, and this difference allows researchers to work with the heterogeneity of language. As the field of language variation and change better understands how the linguistic and nonlinguistic systems interact, variationist researchers will further improve the descriptions and explanations of the resulting variation.

4 The Locus of Language Variation

Justifying the linguistic variable was one of the points of contention between variationists and other linguists in the 1960s and 1970s. The source or locus of language variation in the mind has been another. Weinreich, Labov, and Herzog (1968) established one of the earliest and most resounding arguments for language variation to be seen as inherent variability in a grammatical system. In doing so, they denounced the homogenous view of language, where systematic processes were only categorical, a view touted by their contemporary generative grammarians on language data. Chambers (2003: 12) characterizes that early generative assumption as the axiom of categoricity: "the simplifying assumption that data for linguistic analysis must be regularized to eliminate real-world variability." Weinreich et al. argued that such a view prevented advances in diachronic linguistics and that the homogenous approach presented a "needlessly unrealistic" model for synchronic linguistics. They proffered instead that:

> The key to a rational conception of language change – indeed, of language itself – is the possibility of describing orderly differentiation in a language serving a community. We will argue that nativelike command of heterogeneous structures is not a matter of multidialectalism or "mere" performance, but is part of unilingual linguistic competence. (Weinreich et al. 1968: 101)

In Labov's more recent publications, as in his foundational works, he posits that the individual speaker's grammar, as learned from the speech community, contains the mechanisms of variation, which interacts with linguistic constraints and social constraints.

Fasold and Preston (2007) provide perspective on and comparison with the variationist view of inherent variability and the nonvariationist, generativist view of competing grammars, where the variation is not within a single grammar but results from two or more grammars "competing" to produce the output forms. In describing how these approaches have developed in the different linguistic fields, Fasold and Preston provide three levels of sociolinguistic work

which explain the locus of language variation. In Level I, variationists correlate linguistic and social facts, noting how the presence of one fact increases or decreases in relation to another (e.g., whether lesbians have a more raised /æ/ compared to straight women). Although such correlations become important tools for variationists, at the beginning of the sociolinguistic variationist research effort, Labov (1966: v) issued an early warning about being satisfied with such descriptive correlations:

> In the past few years, there has been considerable programmatic discussions of *sociolinguistics* at various meetings and symposia. If this term refers to the use of data from the speech community to solve problems of linguistic theory, then I would agree that it applies to the research described here. But *sociolinguistics* is more frequently used to suggest a new interdisciplinary field – the comprehensive description of the relations of language and society. This seems to me an unfortunate notion, foreshadowing a long series of purely descriptive studies with little bearing on the central theoretical problems of linguistics or of sociology.

Fasold and Preston (2007: 45) agree that such purely descriptive studies benefit the kind of ethnographic sociolinguistics fostered by Dell Hymes (Johnstone and Marcellino 2011). However, most scholars work beyond this level, using such correlations to develop more complex explanations for linguistics patterns.

In Level II variationist studies, researchers explore the impact of one linguistic factor upon another. In other words, how the components of the mental grammar interact with each other to create the inherent variability observed in language. For example, Weldon (2007) examines five past tense copula variants to assess negation patterns and how subject type, person–number context, and grammatical environment affect the distribution of the variants (e.g., *Well she didn't went no further than that*). Many variationist studies of the last four decades combine Level I and Level II studies.

Fasold and Preston (2007: 61) posit the examination of diachronic variation as Level III sociolinguistic variationist work. Such linguistic studies follow in the footsteps of Labov (1963), who first applied the synchronic ordering of different generations to diachronic problems in order to draw from the apparent-time construct (Bayley 2002). The apparent-time construct assumes that the vernacular of one generation was primarily learned in the adolescence of that generation. Thus, looking at the language variation patterns of different generations allows researchers to study the vernacular of progressively earlier time periods. Variationists have found that some language change does occur during speakers' lifetimes (e.g., Sankoff and Blondeau 2007), but the assumptions of the apparent-time construct continue to hold up under scrutiny and allow for a wealth of explanation. Fasold and Preston (2007) cite Tagliamonte (1998) who investigates different generations as part of the analysis of past tense *be*. Tagliamonte (1998) found that in plural existential constructions (e.g., *There was three of them*), the younger speakers had higher rates of *was* than older speakers. Given this trend, an assessment using the apparent-time construct allows Tagliamonte to predict that the rate of *was* in plural existential construction is increasing over time. Hazen (2014: 13) finds the opposite trend for speakers in the West Virginia region of Appalachia, where the rate of leveled *was* with plural existential subjects fell dramatically from oldest speakers to youngest speakers; hence it appears that this existential construction is decreasing for these speakers. The apparent-time construct allows researchers to make these kinds of reasoned conjectures.

Influences on the linguistic system can also come from several sources. In Vanuatu, Meyerhoff (2000) examines Bislama, a Melanesian creole, in which verbal morphology is influenced by the lexified language, the substrate language, and Universal Grammar. This complex interplay of assumptions, levels, and data yields thorough explanations of language variation patterns. Fasold and Preston recognize that with Level III, both of the other two levels must be used to build the quantitative details necessary to conduct a well-articulated argument about diachronic variation.

5 Rules and Constraints on Variation

Linguists have generally followed two separate explanations for how language variation gets done by its speakers. Within the mind, language formation can either be processed by rules or shaped by constraints.[2] Rules take linguistic raw materials and do something to change them. The raw material could be phonetic in nature so that perhaps places of articulation may be changed (e.g., /k/ → [t] so that *smirk* may be *smirt* for some Southern US speakers). The change could also be morphological. Finkel and Stump (2007) devise a series of rules to model how the mental grammar derives Hebrew verbs from their roots (e.g., d-b-r 'speak' → *Medaber* 'I speak'; *Dibarti* 'I spoke').

When imagining constraints, the linguistic raw material should be considered to be in the process of change; the constraints shape how that change occurs. For example, in the Walmatjari language of Western Australia, the stress on words is variable, with the number of syllables opening up extra locations. To model how stress is assigned, constraints on how prosodic units are built, such as limiting the shape of phonological feet, filter out the best and worst candidates. In the preferred models, the best candidates are those the speakers produce (Anttila 2002: 217). For Walmatjari, *yápa* 'child' receives first-syllable stress, but the *yutanti* 'sit' can receive first- or second-syllable stress. A properly ordered set of constraints should predict such a predicament. Nagy and Reynolds (1997) develop such a model with an Optimality Theory account of word-final deletion in Faetar, a Francoprovençal dialect of southern Italy. Their model is specifically crafted to have floating constraints with variable rankings.

Scholars of language variation have generally divided rules and constraints into two types: external and internal. The term *internal* refers to the linguistic rules and constraints (i.e., those *inside* the mental grammar). The term *external* refers to the social constraints represented elsewhere in the mind. These are not always the most clear terms, but they are common. Where lexical constraints fall is a matter of debate, as traditional models of the mental grammar have the lexicon outside of it, even though the lexicon itself is clearly a mixture of linguistic structure and social factors. The modules of the mind which evaluate social characteristics are traditionally seen as external to the mental grammar.[3] Internal constraints range across the entire linguistic system and are detailed more fully below. External constraints comprise any social difference which might contribute to language variation patterns, such as age, social class, ethnicity, sex, gender, region, and sexual orientation, as discussed by Chambers (see Chapter 24, this volume). At first glance, such a broad range of social and linguistic factors may seem bewildering or perhaps overly ambitious; however, variationists have learned that synchronic and diachronic variation results from numerous causes and constraints. Bayley (2002: 118) has dubbed this truism the principle of multiple causes, which "…means that it is unlikely that any single contextual factor can explain the variability observed in natural language data." Variationists' insistence on natural data requires patience with complex sociolinguistic patterns.

6 Quantitative Analysis

Bayley (2002) also describes one of the key aspects which renders variationist work distinctive from other areas of linguistics: the principle of quantitative modeling. This principle

> means that we can examine closely the forms that a linguistic variable takes, and note what features of the context co-occur with these forms…. With a large enough set of data, we are able to make statements about the likelihood of co-occurrence of a variable form and any one of the contextual features in which we are interested. (2002: 118)

The data sets in variationist studies have fluctuated widely over the years, from small to exceptionally large, and with the regular use of corpus linguistics and sociophonetics, the average size of data sets has grown over the decades. Kendall (2013) describes the innovations in data structure, data mining, and the wide array of ways variationists can implement data-driven studies.

Quantitative analysis can take many forms, but variationists from the 1960s through the 2000s relied predominantly on percentages, crosstabulations, and multivariate analysis (Guy 2014; Walker 2011).[4]

Percentages are the most direct means of assessing sociolinguistic variation. As Tagliamonte (2007: 190) argues:

> Quantitative analysis involves an examination of individual instances of linguistic form in the context of the grammar from which they come. Thus, quantitative analysis is not so much interested in individual occurrences of linguistic features (i.e., tokens), but the recurrent choices an individual makes in the course of linguistic performance (i.e., patterns of occurrence) (Poplack and Tagliamonte 2001: 89). These choices are taken to represent the (underlying) variable grammar of the speaker as well as the speech community to which she belongs.

Percentages reveal those patterns of occurrence and thus give insights about the model of the grammar. In an analysis of changes in Canadian English, Chambers (2002: 364) displays percentages to show differences between two surveys, one from 1920 and one from 1980. For a variable such as the alternation between the voiceless [ʍ] and the voiced [w] in the words *which* and *whine*, the 1920 respondents had rates for [w] of 29 percent in *which* and 31 percent in *whine*; the 1980 respondents had rates of [w] at 89 and 95 percent for the same words, respectively. Before the adoption of acoustic methods and sociophonetics, this kind of phonetic symbol analysis of sounds was the norm (see Section 11 below), but in modern studies of sound variation, acoustic analysis is the norm. The traditional procedure for calculating these percentages is to take the number of occurrences of the variant in question and divide it by the total number of possible tokens: If there were 100 possible instances of *whine* in the data set, 31 of them would be pronounced with a [w] by the 1920 speakers. Possible tokens include all variants of the variable; here that would be all potential instances of word initial [w] and [ʍ]. This practice follows Labov's principle of accountability (Labov 1969), which prevents researchers from fudging their data by cherry-picking the tokens that display categorical tendencies and ignoring the "bothersome" variable elements.

Crosstabulations are a simple yet powerful setup of percentages. Instead of giving a unified social category for percentages, the crosstabulation data display provides a set of subdivisions of that social category. Tagliamonte (2006: 183) demonstrates crosstabulation with final-consonant deletion, showing how education groups break down against age divisions. In her example, the rate of deletion of the final consonant for the higher education group is 27 percent, as it is for the lower education group. From such an analysis, education would appear to have no effect on final-consonant deletion. Yet when the educational groups are subdivided into age groups, striking differences arise. The second oldest age group shows 31 percent deletion for the lower education division, but 14 percent for the higher education division. Why this division for this age group arises requires careful social explanation, but without the crosstabulation of social categories to detect the differences, such an explanation would never be developed.

Multivariate analysis was first applied to language variation in the early 1970s. Cedergren (1973) and Cedergren and Sankoff (1974) developed Labov's (1969) concept of the variable rule and applied it to various sets of data, including Cedergren's work with syllable-final /r/ in the Panamanian Spanish of 79 residents of Panama City. Multivariate analysis allows for several dependent variables to be evaluated together against a large number of sets of influencing factors. For example, Cedergren's (1973) study considered the syllable position and morphemic status of the /r/ along with the phonetic nature of the following segment.

From about the mid-1970s until 2010, to publish on language variation and change usually required multivariate analysis in VARBRUL, first used to support the now defunct concept of the variable rule (e.g., Labov 1969; Fasold 1991). Since 2013, researchers are expected to use mixed-effects models (Drager and Hay 2012; Tagliamonte and Baayen 2012) and other advanced statistical techniques (e.g., Baayen 2008; Gries 2013; K. Johnson 2011; D. Johnson 2008; Paolillo 2002).

Drager and Hay (2012) argue for the use of mixed-effects models to better account for random and fixed predicting variables. For example, in any study that records information on the token words and the speakers, these two factors can have wide-ranging impacts, but the mixed-effects model can handle such fluctuations by considering word and speaker to vary randomly. They further illustrate how part of those models, the random effect intercepts, are useful for the study of speaker style and identity as well as normalizing for vocal tract size. Tagliamonte and Baayen (2012) demonstrate how variation between *was* and *were* in the past-tense *be* paradigm, traditionally analyzed with VARBRUL, can be better investigated with mixed-effects models, random forests, and conditional inference trees.

Advanced statistical techniques have also been implemented with traditional realms of language variation, such as dialectology. Although dialectology is over a century old, scholars are still making innovations and improvements. Grieve, Speelman, and Geeraerts (2011: 3) write that "Adopting a statistical approach not only avoids the identification of spurious regional patterns, but also allows for the identification of regional patterns that may have gone unnoticed in a traditional analysis." They have developed their methods with a set of three statistical techniques: spatial autocorrelation, factor analysis, and cluster analysis. Their statistical work avoids subjective analysis but also is an alternative to dialectometry, which aggregates variables through statistical techniques (Nerbonne and Kretzchmar 2006).

All quantitative analysis provides inherent satisfaction to language variation researchers. It yields order, and in the scientific quest to explain human language, it presents ordered categories. However, variationists must move beyond the climatic feel of well-ordered quantitative results and resist the temptation to let them be the end of the research. Rickford (2002a) laments that the satisfaction of descriptive variationist accounts has preempted the explanation of language variation. Rickford (2002a: 160) writes:

> To my mind, a major flaw in the literature on linguistic variation is the tendency to be satisfied with the data orderings provided by our heuristic tools (frequencies, variable rule programs, implicational scales), without seeking to explain them in linguistic (or social) terms…. Whether our scales are for the variable use of linguistic features, for intuitions, or for patterns of language acquisition, we should not be satisfied to locate descriptive regularities without attempting to explain them.

For the variationist, quantitative analysis of language phenomena is a tool to be used for a large project but not the end in itself.

7 Lexicon

Lexical variation is the realm of variationist analysis with the longest pedigree of scholarship. Both in Europe and the United States, dialectologists have produced large-scale atlases for over a century. Jules Gilliéron and Edmond Edmont published the linguistic atlas of France between 1902 and 1910 based on a questionnaire of 2,000 words (Gilliéron and Edmont 1902). The work of Kurath (e.g., Kurath 1949; Kurath and McDavid 1961) provides an excellent example of dialect geography in the United States; this body of work affected future variationists by hypothesizing boundaries of dialect variation. Kretzschmar and Schneider (1996) update the methodological efforts of Kurath and McDavid by introducing both quantification and computer analysis. In focusing on the Linguistic Atlas of the Middle and South Atlantic States,[5] they provide an introduction to the quantitative analysis of linguistic survey data. Their efforts clearly present the

challenges for modern scholars of language variation. To accurately accommodate large-scale statistical analysis of survey data, the data handling processes often must be automated.

Lexical variation can be defined as the alternation between various lexical items for the same object or idea. In the USA, the most discussed lexical variation is that for a carbonated, sweetened beverage: pop, soda, or coke (Von Schneidemesser 1996). Given the nature of lexical items, where the relation between the form of the word and the meaning of the word is essentially arbitrary, every language contains regular lexical variation. Semantic classifications have been proposed for types of lexical variation (Geeraerts, Grondelaers, and Bakema 1994). Variationists have in general avoided lexical variation as a specific focus in recent decades, instead using it primarily as a means of communicating with the public and satisfying public interest about the topic. The popularity of lexical dialect maps attests to their importance for scholars of language variation: The Harvard Dialect Survey conducted by Bert Vaux and Scott Golder became a topic of major interest when Josh Katz developed new maps of the lexical and pronunciation data (Katz 2016).

Lexical categories have also been shown as constraints in the variation of sign languages, demonstrating the crossmodal effects of basic grammatical groupings. Lucas and Bayley (2005) analyzed American Sign Language data, including the variables of DEAF,[6] signs like KNOW, and 1-handshape signs, with both phonological and grammatical context, such as noun, adjective, and predicate. Across a large set of data, the grammatical factors consistently were found to be the most significant in accounting for the variance of the data.

Additionally, lexical variationist studies do provide quality cultural analysis of specific communities. For example, Childs and Mallinson (2006: 3) argue that "lexical items may serve a significant indexical function in the social construction of ethnicity." They explore how friends in a small African American community in the mountains of North Carolina employ culturally loaded lexical items to establish their identity within and beyond the community. From such analysis, it is evident that lexical variation should be a regular realm of sociolinguistic variationist scholarship.

8 Morphology

Morphology has played a role in language variation research since the start of modern variationist studies, but its role is rarely represented as a unified system analyzed in and by itself. The (ING) studies mentioned in Section 3, as well as the negation studies, are in some respects morphological studies. Labov's (1972a) article on negation was specifically a morphosyntactic paper. For many variables in other areas of language variation, morphology plays a secondary yet important role in influencing the frequency and quality of language variation. For final-stop deletion (see Chapter 24, this volume; Guy 1991a; Temple 2009), most studies have found that whether a consonant or a vowel follows the stop stringently restrict the range of deletion, so that the /t/ of *last field* will be deleted more than the /t/ of *last owner*. Most studies have also found that morphology has an effect on the rate of variation (Tagliamonte and Temple 2005). For example, the forms *past* and *passed* may both be pronounced as [pæst], but for most communities, the /t/ of *past* will be deleted more often than the /t/ of *passed*.[7] Such differences between monomorphemic forms and bimorphemic forms are found in English-speaking communities. However, such constraints are themselves open to variation. In Tagliamonte and Temple (2005) and Hazen (2011b), the morphological influence was slight at best; for their speakers, final-stop deletion was predominantly a phonological process. The crosscurrents of influence can be seen when morphology, phonology, and lexical patterns overlap.

Some variationist studies focus exclusively on the morphological system. For example, Sanchez (2008) examines the system of verbal morphology in the context of structural borrowing. Sanchez reveals that structural linguistic factors, not social ones, determine how borrowing

plays out through quantitative assessment of a creole variety, Papiamentu, which is in contact with Spanish, English, and Dutch. Numerous variationist studies include morphological and morphosyntactic factors as types of influence. For example, in a study of French-language chat discourse, van Compernolle (2008) finds subject type to be an important factor influencing the presence or absence of *ne* as a negative particle.

On the traditional morphological front, variation is a constant point of interest, although variation most often implies variation between languages. Spencer and Zwicky (2001) devote an entire section (V) to the range of morphological variations for different language families. However, these descriptions would not be considered "variationist," since they do not take up quantitative variation or, for the most part, move beyond the axiom of categoricity (Chambers 2002; Tagliamonte 2012).

The question of how to analyze morphological variation has been raised several times since the 1960s. For Buchstaller (2009), the key element is correctly delimiting the envelope of variation for the morphological forms. As she writes:

> The denominator over which we normalise our data is especially crucial. If we decide to group our variants into a linguistic variable, we need to find a principled way to operationalise the set of alternating variants…it is important to close the set that defines the variable, which means to include all the possible variants and thus to provide a stable denominator for the quantitative analysis. (Buchstaller 2009, p. 1027)

In an overview on morphological variation in sociolinguistics, Davydova (2014) identifies potential influencing factors, the effects of other linguistic levels, and the methodological choices facing researchers of morphological variation. As Davydova illustrates, morphological variation is well suited for analysis of diachronic variation and the nature of any particular language's lexicon. Morphological variation develops in numerous systems across languages, and two areas where it has been extensively studied are intensifiers (e.g., Ito and Tagliamonte 2003) and quotatives (e.g., Buchstaller 2013).

An excellent example of disentangling the form and function qualities of morphological variation comes up in a study where the same form is used both as an intensifier and a quotative. Rickford, Wasow, Zwicky, and Buchstaller (2007) conduct a synchronic and diachronic investigation of the lexeme *all* in its duties as an intensifier and quotative. Through quantitative variationist analysis, they illustrate that the intensifier *all* is not a morphological innovation, but instead operates within a wider range of syntactic environments. They further demonstrate that *all*'s quotative role was an innovation in California in the 1980s. Their study shows how morphological variation can be assiduously tracked over linguistic and time dimensions.

Quotatives themselves have also played a major role in research studies of language variation. Buchstaller (2013) provides an overview of this research into morphological variation, as quotatives have been analyzed from discourse analysis, sociolinguistic, typological, construction grammar, and more formal approaches. The rich utility of morphological variation reveals not only linguistic insights but also social patterns. D'Arcy (2010) studied how Maori and Pakeha English New Zealanders employ quotatives. These different ethnic groups build their dialogues differently, especially in regard to quotatives. Linguistic qualities such as verb tense, time reference, and the form of the quotative are used to mark social boundaries. As illustrated by D'Arcy's study, morphological variation is a rich realm for linguistic and sociolinguistic research.

9 Syntax

With the turn away from structuralism and toward generative grammar in the 1960s, the study of syntax became central to formal linguistics. Yet even though the study of language variation

and change grew rapidly at the same time, syntax became the least featured linguistic system in its repertoire. Some scholars did conduct assessments of syntactic variation, but these works were minority members in the field. For example, Labov (1972a) argued for a variable, rule-based syntactic system to explain the processes of three kinds of negative transfer in Black English: negative attraction to subject *any* (e.g., **anybody doesn't go → nobody goes*), negative postposing to indeterminates (e.g., *she doesn't like anything → she likes nothing*), and negative concord (e.g., *they don't like anything → they don't like nothing*). Sells, Rickford, and Wasow (1996) also build grammatical models for negative inversion in AAVE, but they base their approach on Optimality Theory to capture the variable nature of the process. Heidelberger (1978) analyzed the acquisition of German syntax by foreign migrant workers. On the diachronic front, Kroch (1989a, 1989b) has proposed, from his investigations of the rise of periphrastic *do* in English, the constant-rate hypothesis, where syntactic change progresses at the same rate in all contexts.

In recent years, however, scholars have developed dedicated syntactic projects (Green 2007). Often, these projects directly search for what has been dubbed *microvariation* of parameters. Traditional syntacticians investigate differences and similarities between nonmutually intelligible varieties (Baker 2002) through parameters of the Universal Grammar, such as the headedness parameter. In contrast, variationist syntactic scholars usually investigate differences between mutually intelligible varieties predominantly modeling this intralanguage variation with microparametric settings, ranked constraints in Optimality Theory, or competing (sub)grammars (Fasold and Preston 2007; Green 2007). For English in Belfast, Ireland, Henry (1995, 2002) and Wilson and Henry (1998) consider parameter theory itself as a constrained and formalized casting of language variation. Echoing the debates of earlier decades, in arguing against a "competing grammars" theory of syntactic variation, Henry (2002: 272) disputes Kroch's (1994) conclusions about the possibilities for acquiring (or not) periphrastic *do* into the mental grammar. In addition, Henry's own syntactic research (1995) highlights variability in a range of syntactic structures synchronically, although not evenly distributed throughout the speech community.

Even language variation patterns in sociolinguistic variationist research are now being examined from psycholinguistic and other linguistic perspectives. The Northern Concord rule has drawn interest from several fields (e.g., "All of our brothers and sisters *lives* here") (Tagliamonte 1998; Smith and Tagliamonte 1998; Wolfram and Christian 1976; Wolfram, Hazen, and Schilling-Estes 1999). Bock et al. (2006) argue for a psycholinguistic theory that explains verb and pronoun agreement differences between British and American English sentence subjects that have collective head nouns within a parallel architecture of lexical and syntactic formulation.

European studies of nonstandard dialect syntax have a long history, and the work continues in the twenty-first century. Barbiers, Cornips, and van der Kleij (2002) focus on such large-scale dialect syntax projects in Europe, specifically syntactic microvariation of nonstandard dialects. Their topics include such traditional syntactic realms as negation, left periphery, complementizers, and agreement. Cornips and Corrigan (2005) draw together authors who examine the cross-section of syntactic variation, variationist methodology, and social factors. They (2005: 1) explicitly designed their work to "contain treatments incorporating the analysis of external factors into accounts focusing on the internal linguistic conditioning of syntactic variation and change cross-linguistically." Variationist methodologies have also been successfully employed for traditional realms such as clitic placement. Pappas (2004) developed a variation analysis of the placement of weak clitic pronouns in Medieval Greek. His primary focuses were the factors inducing this variation and how it came to be resolved in later (Modern) Greek.

Within a different area of syntactic assessment, Martin and Wolfram (1998) provide a typology/description of African American English sentences. In another extensive descriptive account, Green provides (2002: Chapters 2 and 3) an explanation of verbal markers and morphosyntactic properties in African American English. These kinds of descriptions are not quantitative

themselves, but quantitative accounts are not possible without these thorough explanations of the semantic, morphophonological, and syntactic patterns of African American English.

Although the scholarship of syntactic dialect variation is by no means overpopulated, its scholars in recent years have developed a strong enough research program to allow for assessments of the field, three of which are noted here. Henry (2002) provides assessments of variationist work in syntactic study. She compares variationist studies with Chomskyan syntactic work, reviewing the nature of data in the two areas, the synchronic/diachronic split, and the nature of variability and frequency in syntax. She also makes an argument for systematic variability in the "core" syntax. Adger and Trousdale (2007) draw forth the theoretical implications from an assessment of syntactic work in English dialect study. They address issues of competence and performance (see Chapter 24, this volume) in addition to the intersections of traditional dialectology, syntactic study, and variationist sociolinguistics. From this review, Adger and Trousdale address the trouble with syntactic variables and examine how such variables are handled by modern theoretical models. Kortmann (2006) delivers an assessment of English syntactic variation from a global perspective. He surveys the most interesting possibilities of language variation in English, covering the noun phrase (e.g., *us* as a possessive marker: *We like us town*), verb phrase (e.g., double modals: *We might could go*), negation (e.g., multiple negation: *She won't do no harm*), agreement (e.g., zero marking: *It pop up every time*), and subordination (e.g., relative particle *what*: *This is the one what painted my house*). Kortmann (2006) also characterizes the processes involved in the propagation of variation and then assesses the varieties from a typological perspective.

Some variationist work views morphology and syntax as components of an integrated system. For example, Adamson (1988) draws comparisons and conclusions from crossproducts and VARBRUL for morphological variation, at the edges of syntax and semantics. Adamson's work dovetails with Bybee's research program, which investigates the prototypical structures that lexically support morphological categories (e.g., Bybee 2007). Adamson (1988) argues that such syntactic concepts as the constructs found in construction grammar greatly resemble variable rules (minus the ranking of factors) and that variationist methods could be used to assess their relative strength.

10 Phonology

As Guy (2007) remarks, the study of linguistic variation is often mistakenly perceived to deal exclusively with phonological phenomena. Although variationists have investigated other realms of language, phonology is the component of the mental grammar with the most fine-grained and productive associations with the social system.

Wolfram has employed phonological linguistic variables to assess social relations for four decades, and much of his work illustrates the kinds of phonological variables traditionally studied. In Wolfram (1969), phonological variables are used to show social stratifications by ethnicity and sex, including final-stop deletion, /θ/ → /f/, devoicing of syllable-final voiced stops, and R-vocalization. The phonological process /θ/ → /f/ is one that is found in varieties of English around the world (Schneider 2004: 1123). It involves the voiceless interdental fricative /θ/ changing its place of articulation to a labiodental /f/ in word-medial or word-final position. At times it can also be word initial; for example, David Beckham in some interviews talks about *"free"* (*three*) *goals* and kicking the ball *frough the middle*. In the US South, this dialect feature can be completely unstigmatized, but in the Northern city of Detroit, Wolfram (1969) found it to be a highly stigmatized feature. For devoicing, although final stops are devoiced as a categorical rule in some languages, such as varieties of German, they can be variably devoiced in some varieties of AAVE. R-vocalization is the process of reducing the constriction on a rhotic /ɹ/.[8] This loss of constriction transforms the sonorant /ɹ/ into a vowel within the rhyme of the syllable. R-vocalization does

not normally occur in the onset of a syllable. Most prominent in British dialects, it is a feature of Southern US English and AAVE.

Wolfram (2003) reexamines the development of African American English by synchronic studies of relatively isolated African American communities, and phonological variables play a crucial role in his argument. His premise is that African American communities earlier in the twentieth century primarily demonstrated phonological convergence with their local European American counterparts. For example, he finds vowel systems for older speakers to be vastly more comparable than those for younger speakers (based on research in Wolfram and Thomas 2002). Importantly, however, these isolated African American communities maintained creole-like traits, such as final-stop deletion before vowels, which were more similar to other African American communities with which they had no contact.

Sign languages have been extremely beneficial for widening the scope of phonology and informing phonological theory (Brentari 1995). Sign languages have also been the focus of variationists in their efforts to map the nature of phonological variation. Lucas (2007) is one of several overviews of the variationist sociolinguistic investigation of sign languages. In Lucas's studies of phonological variation, factors considered significant include the preceding location of the sign (body or head), whether or not there was following contact, and the shape of the hand (e.g., the 1-handshape with fingers closed or open). With spoken languages, divisions are often made between paradigmatic organization and syntagmatic organization: A paradigm might consist of a pattern for how to conjugate of verb, whereas a syntagmatic pattern could be the ordering in a phrase. Sign languages have numerous simultaneous qualities, and variation in simultaneity is one feature open for exploration (Brentari 1995).

In Anttila's (2002) assessment of the innovations of phonological variationist studies, he tackles how Optimality Theory has benefited variationist understanding and how variationist models have changed Optimality Theory. In evaluating different kinds of grammars, Anttila judges whether the originally categorical grammars allow for modifications to a quantitative paradigm. He is careful to weigh the different grammars for their ability to handle quantitative variation but to also ensure that such modifications do not destroy the theoretical generalizability of the Optimality Theory grammars. Anttila (2002: 236) concludes that Optimality Theory has given "hope that generative phonology is beginning to answer some of the questions raised by variationist linguists."

Guy (2007) reviews several of the most common phonological constraints on variation, including stress and the nature of following segments. Word stress and varying degrees of phrasal stress are found to influence phonological variation in practically every language. For example, in numerous languages, unstressed vowels are more likely to be centralized to schwa [ə]. In coordination with influence from syllable structure, the nature of the following segment, or lack thereof, also affects phonological variation, as was noted for final-stop deletion (see Chambers, this volume). The articulatory and acoustic phonetic qualities of following segments also affect variation. For example, Guy (2007) found that in Brazilian Portuguese, higher rates of deletion of final sibilants (e.g., /s/) occurred before voiced consonants and alveolar consonants. Careful examination reveals how such variation informs phonological theories: Final sibilants in Brazilian Portuguese take up the same voicing as their following segments and voiced fricatives are universally more marked and therefore open to deletion; the higher rate of deletion before alveolars conforms to the Obligatory Contour Principle, which claims that adjacent, identical elements are dispreferred. The implication is that such phonological tendencies affect categorical and variable linguistic patterns (Guy and Boberg 1997).

Articulatory effects are found for variable processes in languages around the world. In a study of the Arabic of Cairo, Haeri (1997) examines palatalization socially and linguistically. Palatalization is the process whereby consonants pick up a more palatal place of articulation, most often through influence from a following high front vowel (e.g., /naadi/ 'club' → [naadʒi]). For Cairene Arabic, palatalization distinguishes the speech of women and men, with women having

more frequent and advanced palatalization. Haeri (1997) illustrates how phonological variables reveal sociolinguistic trends synchronically and diachronically.

Foulkes (2006) outlines the causes and effects of phonological variation, predominantly from English but with an ear to other languages. Foulkes also notes the importance of the study of speech perception, while reporting from the vastly more developed field of speech production. After reviewing the biological constraints on phonological variation, Foulkes (2006) argues for the inclusion in the study of phonological variation contextual constraints such as coarticulation, which is the effect of one sound on another (e.g., the rounded second vowel in *see-saw* usually affects the second /s/ so that it is articulated with rounded lips). In addition, Foulkes details the interaction of phonology with other linguistic systems and highlights the constraints of social variation, such as geography, social class, and age. He stresses that findings about phonological variation are important for the development of models of phonological theory, as advanced and realistic models must account for such variation.

11 Phonetics

The study of variable sound patterns through acoustic phonetics has been the fastest growing area in the study of language variation in recent years. This branch of linguistics is now called sociophonetics (Thomas 2011, 2014). At its base, it combines acoustic analysis with statistical analysis. Both of these types of analysis are applied to research questions about patterns of sounds and the social qualities that influence them. A basic distinction between phonology and sociophonetics is the "packaging" and concept of the sounds they study. A phonological approach would ask if, how, and why the vowel /æ/ in a word like *bad* became similar to the vowel /ɛ/ in *bed*. A sociophonetic approach would use computer software to quantitatively measure many spoken instances of *bad* in order to measure qualities such as duration and formants and then would compare them to the same qualities of *bed* pronunciations. Almost all new studies of language variation in sound patterns are sociophonetic studies.

Although the name *sociophonetics* was new at the turn of the century, the use of acoustic methods is not new in sociolinguistics. Labov's (1963) foundational study of Martha Vineyard's vowels employed acoustic techniques to support the auditory vowel assessments, and Labov, Yaeger, and Steiner (1972) used them as part of the study of vowel systems, one of the most popular topics in the 50-year history of sociolinguistics. Researchers have conducted studies of variation in consonants, but these are greatly outnumbered by those examining vowels. As an example of consonant variation, devoicing in Wolfram (1969) was analyzed through auditory assessment. With phonetic analysis, the concept of binary devoicing (voiced or voiceless) is overly simple and requires refinement. Purnell, Salmons, Tepeli, and Mercer (2005) argue for an expanded understanding of devoicing as a quantitatively and qualitatively variable process. Instead of relying on a binary category of glottal pulsing, they demonstrate the need to assess the degree of glottal pulsing, the duration of the segment in question, the duration of the preceding segment, the pitch and formants of the preceding vowel, and other acoustic qualities. As the study of variation has advanced, the linguistic assessment of the object of study has become more complex.

The task for language variation scholars has been to find order in the complexity, and with this goal in mind, variationists have performed highly detailed and productive studies of vowel systems. One such work with ties to dialectology is Labov, Ash, and Boberg's *The Atlas of North American English* (2006). Their atlas is based on telephone data from across North America elicited through reading passages, word lists, and more casual interviews. Their team of researchers acoustically analyzed the vowel systems of 439 of the speakers interviewed to produce geographic maps illustrating patterns of synchronic variation on a national scale.

Many vowel studies investigate and compare the vowel space of a speaker or community to that of others. For example, Thomas (2001) performs acoustic analysis on hundreds of speakers

from the US North, Southeast, and South-central states, while also comparing across four ethnic groups. In making these comparisons, researchers highlight shifts in the location of vowels in the vowel space. Those locations are produced quantitatively through acoustic analysis software. Because of this type of fine-grained data and the relatively quick changes vowels undergo, the variationist study of diachronic variation often focuses on vowel movements. For example, Baranowski (2007) examined vowel mergers, vowel shifts, and postvocalic /r/ variation in Charleston, South Carolina. Through these phonological variables, Baranowski reveals that younger generations have replaced traditional patterns with innovations and regional norms. For example, the merger between /ihr/ and /ehr/, as in *fear* and *fair*, was a traditional feature of the city's dialect, but in connection with the introduction of nonvocalized /r/, the merger is reversed in younger generations. Baranowski (2007) illustrates the interaction of linguistic patterns and their combined effects on language variation.

The patterned shifts in vowel systems are often named as coherent processes: for example the Great Vowel Shift, the Northern Cities Shift, or the Southern Vowel Shift (Labov 1994). Some communities push vowel shifts forward to completion, while others retreat from them, depending on sociolinguistic factors. Dodsworth and Kohn (2012) studied the vowel systems of speakers from Raleigh, NC, and found a retreat from the Southern Vowel Shift, a pattern generally advancing in rural areas and certain social groups in the US South (Fridland 2001). Their study also illustrated the statistical innovations used in modern vowel studies, including analysis and visualization through generalized additive models, scatterplots, linear mixed models, and box and whisker plots. This kind of multipronged statistical analysis sets the standard for analysis of vowel systems in the future.

Perception also plays an important role in the study of phonetic variation, and increasingly, researchers are including perception studies with production studies. For example, Kendall and Fridland (2012) investigated how speakers from three US regions performed on a vowel identification task for a continuum between /e/ and /ɛ/. They found that region plays a significant role in mediating perception, as does individual participation in vowel changes, such as with the Southern Vowel Shift speakers.

As can be seen in this small sample of vowel studies, the research methods in language variation are evolving and expanding. A good example of a new method in the study of diachronic variation of vowel systems comes from Stanford and Kenny's (2013) presentation of the first agent-based computer simulation of vowel chain shifts across large communities. Their computer models provide for different interpretations of why vowel shifts play out in certain ways for certain communities. This kind of research tool augments sociolinguistic data collection and acoustic analysis and provides another means of proffering explanations for community-specific changes and changes in the human language system as a whole.

12 Community Outreach

One of the most well-established findings of modern linguistics is that language variation is natural. Diversity is the norm for human language. In the name of this basic premise, variationists have for decades been committed to community outreach and the desire to change public opinion about stigmatized dialects. This outreach has most often taken the form of educational initiatives (for a history of these, see Hazen 2017), though there have been other kinds of successful efforts (see Charity 2008 for an overview). Labov's early work in Harlem on AAVE resulted from the desire to correct educational misconceptions which claimed AAVE was a deficient form of language. Labov showed in both extensive academic work (e.g., Labov, Cohen, Robbins, and Lewis 1968; Labov 1969) and public writings (e.g., Labov 1972b) that AAVE was a rule-governed, logical dialect, despite the heavy stigma laid against it. The Rickford family have developed public materials and educational materials for AAVE and Guyanese Creole; in addition, they represented the variationist sociolinguistic community well in the Ebonics debate (e.g., A. Rickford

1999; A. Rickford, Sweetland, and J. Rickford 2004; J. Rickford 1997; J. Rickford 2002b; J. Rickford and A. Rickford 1995; J. Rickford and R. Rickford 2000). In the same vein, the multi-generational team of Reaser, Adger, Wolfram, and Christian (2017) have developed numerous educational programs about language variation workshops over the last 30 years, providing workshops and curriculums for teacher education, including a chapter on dialect history and diversity in a state-mandated 8th grade textbook in North Carolina.

Several in-depth studies and teacher-oriented publications have focused on the ways in which variationist analysis can assist educational goals: Charity (2005) reports findings from the study of dialect variation of African American children in school settings. Sweetland (2006) studies the development and implementation of language variation teacher training programs. Reaser (2006) examines dialect awareness programs and their effects on teachers' and students' attitudes about language variation. Charity-Hudley and Mallinson (2013) provide a teacher-oriented synthesis of variationist goals for education.

Wolfram and Schilling (2016) review the general principles which have guided sociolinguistic variationists in their attempts to work to the benefit of communities. In general, researchers should seek out ways to use what they have learned to correct misconceptions and provide assistance wherever needed. Variationists who rely on communities for their data should feel proud to help those communities.

13 Conclusion

Variationists examine patterns to better understand how language works. As Warner (2006: 45) writes about variation and grammatical change: "A crucial assumption here is that particular aspects of the way language is used will reflect aspects of its underlying grammar." This assumption is foundational for variationists. Variationists quantitatively describe linguistic systems. For those descriptions to contribute to the accurate explanation of the wide range of human language, the observable parts of language must reflect the mental machinery.

Although the quest for quantitative, variable description began with sociolinguists examining data from English in the United States, modern scholarship reflects increasing diversity in three areas, all of which benefit the variationist enterprise. First, scholars outside of sociolinguistics have taken up the noncategorical examination of variation patterns. Second, language variation studies have been conducted much more widely around the world. Third, English is now only one of many languages under the focus of variationist methodologies. Regardless of which human language is under study, variation provides opportunities for students of language to examine social and linguistic stories of humanity.

SUGGESTIONS FOR FURTHER READING

Bayley, R. and Lucas, C. (eds) (2007). *Sociolinguistic Variation: Theory, Methods, and Applications*. Cambridge: Cambridge University Press.

Chambers, J. K. and Schilling, N. (eds) (2013). *The Handbook of Language Variation and Change*, 2nd edn. Malden, MA: Wiley Blackwell.

RELEVANT JOURNALS

Language Variation and Change
American Speech
Journal of Sociolinguistics
English Language and Linguistics
Journal of English Linguistics

EMERGING TRENDS AND RESEARCH QUESTIONS

1. Stronger connections are being made between variationist sociolinguists and other disciplines of linguistics. As Jen Hay has shown in numerous works and through her leadership at the New Zealand Institute of Language, Brain, and Behaviour (http://www.nzilbb.canterbury.ac.nz), fields such as phonetics, statistics, and brain studies are excellent ways to explore language variation.
2. More linguists of different stripes are taking up the study of language variation. Use of variationist methodology by scholars not self-identifying as sociolinguists has increased over recent years. Variation is for everyone (e.g., Raymond, Dautricourt, and Hume 2006)!
3. The quantitative drive persists, and the need for quantification in linguistic arguments has grown more important. Although basic statistical techniques are still important (Guy 2014), more advanced analysis, especially with visualization, is becoming the norm (e.g., Gries 2013; Tagliamonte and Baayen 2012). Graduate students specializing in language variation must have a side specialization in statistics.
4. Variationist analysis has advanced enough so that scholars can look across many studies, regions, and languages to find larger sets of patterns. Increasing efforts to describe general principles of language variation should continue in this century. Also, scholars must enthusiastically study languages other than English.
5. In the continuing variationist research program, the quest to model variation in the mind will take on greater importance.

NOTES

1 For decades, the parentheses around the linguistic form indicated that it was the name of the sociolinguistic variable, much the same way slashes indicate a phoneme, such as /æ/. This practice of ( ) around a variable's name is only occasionally applied by researchers of language variation today, as sociophonetics has rendered many of the older phonological practices obsolete.

2 Although in recent years variationists have paid more attention to perception, most variationist studies have examined production (see Thomas 2002).

3 The locus of the lexicon is crucial in asking social questions, as the lexicon is a linguistic unit which stores detailed social information including social evaluations.

4 Implicational scales were also a component of early variation analysis (J. Rickford 2002a).

5 See http://us.english.uga.edu/lamsas/

6 ASL signs are written in English with all capital letters to distinguish them from English words. The word "deaf" is English, but the sign DEAF is ASL.

7 Why this situation occurs is another question: Guy (1991a, 1991b) argues for an lexical phonology explanation, while Bybee (2007) argues for an exemplar model explanation.

8 There are several symbols for the R sound in American <red>. Here the IPA upside-down /ɹ/ is used.

REFERENCES

Adamson, H. D. (1988). *Variation Theory and Second Language Acquisition*. Washington, DC: Georgetown University Press.

Adger, D. and Trousdale, G. (2007). Variation in English syntax: theoretical implications. *English Language and Linguistics*, 11(2), 261–78.

Anttila, A. (2002). Variation and phonological theory. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds), *The Handbook of Language Variation and Change* (pp. 206–43. Oxford: Blackwell.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baker, M. C. (2002). *The Atoms of Language*. New York: Oxford University Press.

Baranowski, M. (2007). *Phonological Variation and Change in the Dialect of Charleston, South Carolina. Publication of the American Dialect Society 92*. Durham, NC: Duke University Press.

Barbiers, S., Cornips, L., and van der Kleij, S. (eds) (2002). *Syntactic Microvariation*. Electronic publication of the Meertens Institute. Amsterdam: Meertens Institute. http://www.meertens.knaw.nl/books/synmic/.

Bayley, R. (2002). The quantitative paradigm. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds), *The Handbook of Language Variation and Change* (pp. 117–41). Oxford: Blackwell.

Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart, and Winston.

Bock, K., Butterfield, S., Cutler, A. , Cutting, J. C., Eberhard, K. M., and Humphreys, K. R. (2006). Number agreement in British and American English: disagreeing to agree collectively. *Language*, 1, 64–113.

Brentari, D. (1995). Sign language phonology: ASL. In J. Goldsmith (ed.), *The Handbook of Phonological Theory* (pp. 615–39). Malden, MA: Blackwell.

Buchstaller, I. (2009). The quantitative analysis of morphosyntactic variation: constructing and quantifying the denominator. *Language and Linguistics Compass*, 3/4, 1010–33.

Buchstaller, I. (2013). *Quotatives: New Trends and Sociolinguistic Implications*. Malden, MA: Wiley Blackwell.

Bybee, J. (2007). *Frequency of Use and the Organization of Language*. New York: Oxford University Press.

Campbell-Kibler, K. (2007). Accent, (ING), and the social logic of listener perceptions. *American Speech*, 82(1), 32–64.

Campbell-Kibler, K. (2008). I'll be the judge of that: diversity in social perceptions of (ING). *Language in Society*, 37(5), 637–59.

Campbell-Kibler, K. (2011). The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change*, 22(3), 423–41.

Cedergren, H. (1973). On the nature of variable constraints. In C-J. N. Bailey and R. W. Shuy (eds), *New Ways of Analyzing Variation in English* (pp. 13–22). Washington, DC: Georgetown University Press.

Cedergren, H. and Sankoff, D. (1974). Variable rules: performance as a statistical reflection of competence. *Language*, 50, 333–55.

Chambers, J. K. (2002). Patterns of variation including change. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds), *The Handbook of Language Variation and Change* (pp. 349–72). Oxford: Blackwell.

Chambers, J. K. (2003). *Sociolinguistic Theory*, 2nd edn. Malden, MA: Wiley Blackwell.

Charity, A. H. (2005). Dialect Variation in School Settings Among African-American Children of Low-Socioeconomic Status. PhD dissertation, University of Pennsylvania.

Charity, A. H. (2008). Linguists as agents for social change. *Language and Linguistics Compass*, 2(5), 923–39. Available online at http://onlinelibrary.wiley.com/doi/10.1111/j.1749-818X.2008.00081.x/full/.

Charity-Hudley, A. and Mallinson, C. (2013). *We Do Language: English Language Variation in the Secondary English Classroom*. New York: Teachers College Press.

Childs, R., and Mallinson, C. (2006). The significance of lexical items in the construction of ethnolinguistic identity: a case study of adolescent spoken and online language. *American Speech*, 81, 3–30.

Cornips, L. and Corrigan, K. P. (eds) (2005). *Syntax and Variation: Reconciling the Biological and the Social*. Philadelphia, PA: John Benjamins.

D'Arcy, A. (2010). Quoting ethnicity: constructing dialogue in Aotearoa. *Journal of Sociolinguistics*, 14(1), 60–88.

Davydova, J. (2014). Morphosyntactic analysis in sociolinguistics. In J. Holmes and K. Hazen (eds), *Research Methods in Sociolinguistics: A Practical Guide* (pp. 149–62). Malden, MA: Wiley Blackwell.

Dodsworth, R. and Kohn, M. (2012). Urban rejection of the vernacular: the SVS undone. *Language Variation and Change*, 24(2), 221–45.

Drager, K. and Hay, J. (2012). Exploiting random intercepts: two case studies in sociophonetics. *Language Variation and Change*, 24(1), 59–78.

Fasold, R. W. (1990). *Sociolinguistics of Language*. Malden, MA: Blackwell.

Fasold, R. W. (1991). The quiet demise of variable rules. *American Speech*, 66, 3–21.

Fasold, R. and Preston, D. (2007). The psycholinguistic unity of inherent variability: Old Occam whips out his razor. In R. Bayley and C. Lucas (eds), *Sociolinguistic Variation* (pp. 45–69). New York: Cambridge University Press.

Finkel, R. and G. Stump. (2007). A default inheritance hierarchy for computing Hebrew verb morphology. *Literary and Linguistic Computing*, 22(2), 117–36.

Fischer, J. L. (1958). Social influence on the choice of a linguistic variant. *Word*, 14, 47–56.

Foulkes, P. (2006). Phonological variation: a global perspective. In B. Aarts and A. McMahon (eds), *The Handbook of English Linguistics* (pp. 625–69). Malden, MA: Blackwell.

Fridland, V. (2001). The social dimension of the Southern Vowel Shift: gender, age, and class. *Journal of Sociolinguistics*, 5, 233–53.

Geeraerts, D., Grondelaers, S., and Bakema, P. (1994). *The Structure of Lexical Variation*. New York: Mouton de Gruyter.

Gilliéron, J. and Edmont, E. (1902–1910). *Atlas linguistique de la France* (35 volumes). Paris: Honoré Champion.

Green, L. J. (2002). *African American English*. New York: Cambridge University Press.

Green, L. J. (2007). Syntactic variation. In R. Bayley and C. Lucas (eds), *Sociolinguistic Variation* (pp. 24–44). New York: Cambridge University Press.

Gries, Stefan Th. (2013). *Statistics for Linguistics with R: A Practical Introduction*. New York: De Gruyter.

Grieve, J., Speelman, D., and Geeraerts, D. (2011). A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change*, 23(2), 193–221.

Guy, G. R. (1991a). Contextual conditioning in variable lexical phonology. *Language Variation and Change*, 3, 223–39.

Guy, G. R. (1991b). Explanation in variable phonology: an exponential model of morphological constraints. *Language Variation and Change*, 3, 1–22.

Guy, G. R. (2007). Variation and phonological theory. In R. Bayley and C. Lucas (eds), *Sociolinguistic Variation* (pp. 5–23). New York: Cambridge University Press.

Guy, G. R. (2014). Words and numbers: statistical analysis in sociolinguistics. In J. Holmes and K. Hazen (eds), *Research Methods in Sociolinguistics: A Practical Guide* (pp. 194–210). Malden, MA: Wiley Blackwell.

Guy, G. R. and Boberg, C. (1997). The obligatory contour principle and sociolinguistic variation. *Language Variation and Change*, 9, 149–64.

Haeri, N. (1997). *The Sociolinguistic Market of Cairo: Gender, Class, and Education* (Library of Arabic Linguistics, Vol 13). London: Kegan Paul.

Hazen, K. (2007). The study of variation in historical perspective. In R. Bayley and C. Lucas (eds), *Sociolinguistic Variation* (pp. 70–89). New York: Cambridge University Press.

Hazen, K. (2008). (ING): A vernacular baseline for English in Appalachia. *American Speech*, 83(2), 116–40.

Hazen, K. (2011a). William Labov: language variation and change. In R. Wodak, B. Johnstone, and P. Kerswill (eds), *Handbook of Sociolinguistics* (pp. 24–39). New York: Sage Publications.

Hazen, K. (2011b). Flying high above the social radar: coronal stop deletion in modern Appalachia. *Language Variation and Change*, 23(1), 105–37.

Hazen, K. (2014). A new role for an ancient variable in Appalachia: paradigm leveling and standardization in West Virginia. *Language Variation and Change*, 26, 1–26.

Hazen, K. (2017). Variationist approaches to language and education. In K. A. King, Y. Lai, and S. May (eds), *Encyclopedia of Language and Education: Research Methods in Language and Education*, 3rd edn. NY: Springer. DOI 10.1007/978-3-319-02329-8_10-1., 1–13

Heidelberger Forschungsprojekt 'Pidgin-Deutsch.' (1978). The acquisition of German syntax by foreign migrant workers. In D. Sankoff (ed.), *Linguistic Variation: Models and Methods* (pp. 1–36). New York: Academic.

Henry, A. (1995). *Belfast English and Standard English: Dialect Variation and Parameter Setting*. New York: Oxford University Press.

Henry, A. (2002). Variation and syntactic theory. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds), *The Handbook of Language Variation and Change* (pp. 267–82). Oxford: Blackwell.

Hock, H. H. and Joseph, B. D. (1996). *Language History, Language Change, and Language Relationship*. New York: Mouton de Gruyter.

Holmes, J. and Hazen, K. (eds) (2014). *Research Methods in Sociolinguistics: A Practical Guide*. Malden, MA: Wiley Blackwell.

Houston, A. (1985). Continuity and Change in English Morphology: The Variable (ING). PhD dissertation, University of Pennsylvania.

Ito, R. and Tagliamonte, S. (2003). Well weird, right dodgy, very strange, really cool: layering and recycling in English intensifiers. *Language in Society*, 32(2), 257–79.

Johnson, D. (2009). Getting off the Goldvarb standard: mixed-effects variable rule analysis with Rbrul. *Language and Linguistics Compass* 3(1), 359–83.

Johnson, K. (2011). *Quantitative Methods in Linguistics*. Malden, MA: John Wiley & Sons.

Johnstone, B. and Marcellino, W. M. (2011). Dell Hymes and the ethnography of communication. In R. Wodak, B. Johnstone, and P. Kerswill (eds), *The SAGE Handbook of Sociolinguistics* (pp. 57–66). London: Sage Publications.

Katz, Josh (2016). *Speaking American: How Y'all, Youse, and You Guys Talk, A Visual Guide*. New York: Houghton Mifflin Harcourt Publishing Company.

Kendall, T. (2013). Data in the study of variation and change. In J. K. Chambers and N. Schilling (eds), *The Handbook of Language Variation and Change* (pp. 38–56). Malden, MA: Wiley Blackwell.

Kendall, T. and Fridland, V. (2012). Variation in perception and production of mid-front vowels in the US Southern Vowel Shift. *Journal of Phonetics*, 40(2), 289–306.

Kortmann, B. (2006). Syntactic variation in English: a global perspective. In B. Aarts and A. McMahon (eds), *The Handbook of English Linguistics* (pp. 603–24). Malden, MA: Blackwell.

Kretzschmar, W. and Schneider, E. (1996). *Introduction to Quantitative Analysis of Linguistic Survey*

Data: An Atlas by the Numbers. Thousand Oaks, CA: Sage Publications.

Kroch, A. (1989a). Function and grammar in the history of English: periphrastic *do*. In R. W. Fasold and D. Schiffrin (eds), *Language Variation and Change* (pp. 133–72). Philadelphia: John Benjamins.

Kroch, A. (1989b). Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1, 199–244.

Kroch, A. (1994). Morphosyntactic variation. In K. Beals, J. Denton, R. Knippen, L. Melnar, H. Suzuki, and E. Zeinfeld (eds), *Papers from the 30th Regional Meeting of the Chicago Linguistic Society, Vol 2: The Parasession on Variation in Linguistic Theory* (pp. 180–201). Chicago, IL: University of Chicago Press.

Kurath, H. (1949). *A Word Geography of the Eastern United States*. Ann Arbor, MI: University of Michigan Press.

Kurath, H. and McDavid, R. I., Jr. (1961). *The Pronunciation of English in the Atlantic States*. Ann Arbor, MI: University of Michigan Press.

Labov, W. (1963). The social motivation of language change. *Word*, 19, 273–309.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, 45, 715–62.

Labov, W. (1972a). Negative attraction and negative concord in English grammar. *Language*, 48, 773–818.

Labov, W. (1972b). Academic ignorance and Black intelligence. *The Atlantic Monthly*, 229, 59–67.

Labov, W. (1989). The child as linguistic historian. *Language Variation and Change*, 1, 85–97.

Labov, W. (1994). *Principles of Linguistic Change, Volume 1: Internal Factors*. Malden, MA: Blackwell.

Labov, W. (2001). *Principles of Linguistic Change, Volume 2: Social Factors*. Malden, MA: Blackwell.

Labov, W. (2006). *The Social Stratification of English in New York City*, 2nd edn. New York: Cambridge University Press.

Labov, W., Ash, S., and Boberg, C. (2006). *Atlas of North American English*. Hague: Mouton de Gruyter.

Labov, W., Cohen, P., Robins, C., and Lewis, J. (1968). A study of the non-standard English of Negro and Puerto Rican speakers in New York City. United States Office of Education Final Report, Research Project 3288.

Labov, W., Yaeger, M., and Steiner, R. (1972). *A Quantitative Study of Sound Change in Progress, Volume 1*. US Regional Survey.

Lucas, C. (2007). Variation and modality. In R. Bayley and C. Lucas (eds), *Sociolinguistic Variation* (pp. 145–61). New York: Cambridge University Press.

Lucas, C. and Bayley, R. (2005). Variation in ASL: the role of grammatical function. *Sign Language Studies*, 6, 38–75.

Martin, S., and Wolfram, W. (1998). The sentence in African-American Vernacular English. In S. Mufwene, J. R. Rickford, G. Bailey, and J. Baugh (eds), *African-American English* (pp. 11–36). New York: Routledge.

Meyerhoff, M. (2000). The emergence of creole subject–verb agreement and the licensing of null subjects. *Language Variation and Change*, 12(2), 203–30.

Nagy, N. and Reynolds. B. (1997). Optimality Theory and variable word-final deletion in Faetar. *Language Variation and Change*, 9(1), 37–55.

Nerbonne, J. and Kretzchmar, W. (2006). Progress in dialectometry: toward explanation. *Literary and Linguistic Computing*, 21, 387–97.

Paolillo, J. C. (2002). *Analyzing Linguistic Variation: Statistical Models and Methods*. Stanford, CA: CSLI Publications.

Pappas, P. (2004). Variation and Morphosyntactic Change in Greek: From Clitics to Affixes. New York: Palgrave Macmillan.

Poplack, S. and Tagliamonte, S. (2001). *African American English in the Diaspora*. Malden, MA: Blackwell.

Purnell, T. C., Salmons, J., Tepeli, D., and Mercer, J. (2005). Structured heterogeneity and change in laryngeal phonetics: upper Midwestern final obstruents. *Journal of English Linguistics*, 33(4), 307–38.

Raymond, W. D., Dautricourt, R., and Hume, E. (2006). Word-internal /t,d/ deletion in spontaneous speech: modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, 18, 55–97.

Reaser, J. (2006). A Teacher's Guide for Dialect Awareness Programs. PhD dissertation, Duke University.

Reaser, J., Adger C. T., Wolfram, W., and Christian, D. (2017). *Dialects at School: Educating Linguistically Diverse Students*. New York: Routledge.

Rickford, A. E. (1999). *I Can Fly: Teaching Reading and Narratives to African American and Other Ethnic Minority Students*. Lanham, MD: University Press of America.

Rickford, A. E., Sweetland, J., and Rickford, J. R. (2004). African American English and other vernaculars in education: a topic-coded bibliography. *Journal of English Linguistics*, 32, 230–320.

Rickford, J. R. (1997). Unequal partnership: Sociolinguistics and the African American speech community. *Language in Society* 26, 161–197.

Rickford, J. R. (2002a). Implicational scales. In J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds), *The Handbook of Language Variation and Change* (pp. 143–67). Oxford: Blackwell.

Rickford, J. R. (2002b). Linguistics, education, and the Ebonics firestorm. In J. E. Alatis, H. E. Hamilton, and A. H. Tan (eds), *Linguistics, Language and the Professions* (pp. 25–45). Georgetown University Round Table on Languages and Linguistics 2000. Washington, DC: Georgetown University Press.

Rickford, J. R. and Rickford, A. (1995). Dialect readers revisited. *Linguistics and Education*, 7, 107–28.

Rickford, J. R. and Rickford, R. J. (2000). *Spoken Soul: The Story of Black English*. New York: John Wiley.

Rickford, J. R., Wasow, T., Zwicky, A., and Buchstaller, I. (2007). Intensive and quotative all: something old, something new. *American Speech*, 82(1), 3–31.

Sanchez, T. (2008). Accountability in morphological borrowing: analyzing a linguistic subsystem as a sociolinguistic variable. *Language Variation and Change*, 20(2), 225–53.

Sankoff, G. (2013). Linguistic outcomes of bilingualism. In J. K. Chambers and N. Schilling (eds), *The Handbook of Language Variation and Change*, 2nd edn (pp. 501–18). Malden, MA: Wiley Blackwell.

Sankoff, G., and Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language*, 83(3), 560–88.

Sapir, E. (1921). *Language, An Introduction to the Study of Speech*. New York: Harcourt, Brace, and Company.

Saussure, F. (1972 [1916]). Cours de linguistique générale. Paris: Payot.

Schneider, E. W. (2004). Global synopsis: phonetic and phonological variation in English worldwide. In E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, and C. Upton (eds), *Handbook of Varieties of English* (pp. 1111–37). Berlin: Mouton de Gruyter.

Sells, P., Rickford, J. R., and Wasow, T. (1996). An optimality approach to variation in negative inversion in AAVE. *Natural Language and Linguistic Theory*, 14, 591–627.

Smith, J. and Tagliamonte, S. (1998). 'We was all thegither … I think we were all thegither': Was regularization in Buckie English. *World Englishes*, 17(2), 105–26.

Spencer, A. and Zwicky, A. M. (2001). *The Handbook of Morphology*. Malden, MA: Wiley Blackwell.

Stanford, J. and Kenny, L. (2013). Revisiting transmission and diffusion: an agent-based model of vowel chain shifts across large communities. *Language Variation and Change*, 25(2), 119–53.

Sweetland, J. (2006). Teaching Writing in the African American Classroom: A Sociolinguistic Approach. PhD dissertation, Stanford University.

Tagliamonte, S. (1998). Was/were variation across the generations: view from the city of York. *Language Variation and Change*, 10(2), 153–92.

Tagliamonte, S. (2003). Someth[in]'s go[ing] on!: Variable *ing* at ground zero. In B. Gunnarsson et al. (eds). Language Variation in Europe: Papers from the Second International Conference on Language Variation in Europe, ICLAVE 2 (390–403). Uppsala, Sweden.

Tagliamonte, S. (2006). *Analysing Sociolinguistic Variation*. New York: Cambridge University Press.

Tagliamonte, S. (2007). Quantitative analysis. In R. Bayley and C. Lucas (eds), *Sociolinguistic Variation* (pp. 190–214). New York: Cambridge University Press.

Tagliamonte, S. (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Malden, MA: Wiley Blackwell.

Tagliamonte, S. and Baayen, R. H. (2012). Models, forests, and trees of York English: was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2), 135–78.

Tagliamonte, S. and Temple, R. (2005). New perspectives on an ol' variable: (t,d) in British English. *Language Variation and Change*, 17(3), 281–302.

Temple, R. (2009). (t,d): the variable status of a variable rule. In O. Parker Jones and E. Payne (eds), Oxford University Working Papers in Linguistics, Philology and Phonetics: Papers in Phonetics and Computational Linguistics. http://www.linguphil.ox.ac.uk/files/uploads/OWP2009.pdf.

Thomas, E. R. (2001). *An Acoustic Analysis of Vowel Variation in New World English*. *Publication of the American Dialect Society 85*. Durham, NC: Duke University Press.

Thomas, E. R. (2002). Sociophonetic applications of speech perception experiments. *American Speech*, 77(2), 115–47.

Thomas, E. R. (2011). *Sociophonetics: An Introduction*. New York: Palgrave Macmillan.

Thomas, E. R. (2014). Phonetic analysis in sociolinguistics. In J. Holmes and K. Hazen (eds), *Research Methods in Sociolinguistics: A Practical Guide* (pp. 119–35). Malden, MA: Wiley Blackwell.

van Compernolle, R. A. (2008). Morphosyntactic and phonological constraints on negative particle variation in French-language chat discourse. *Language Variation and Change*, 20(2), 317–40.

Von Schneidemesser, L. (1996). Soda or Pop? *Journal of English Linguistics*, 24(4), 270–87.

Walker, J. A. (2010). *Variation in Linguistic Systems*. New York: Routledge.

Warner, A. (2006). Variation and the interpretation of change in periphrastic DO. In A. van Kemenade and B. Los (eds), *The Handbook of the History of English Language* (pp. 45–67). Malden, MA: Blackwell.

Weinreich, U., Labov, W., and Herzog, M. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann and Y. Malkiel (eds),

Directions for Historical Linguistics (pp. 95–189). Austin, TX: University of Texas Press .

Weldon, T. (2007). Gullah negation: a variable analysis. *American Speech*, 82(4), 341–66.

Whitney, W. D. (1867). *Language and the Study of Language*. New York: Scribner.

Wilson, J. and Henry, A. (1998). Parameter setting in a socially realistic linguistics. *Language in Society*, 27, 1–21.

Wodak, R., Johnstone, B., and Kerswill, P. (eds) (2011). *The SAGE Handbook of Sociolinguistics*. London: Sage Publications.

Wolfram, W. (1969). *A Sociolinguistic Study of Detroit Negro Speech*. Washington, DC: Center for Applied Linguistics.

Wolfram, W. (1993). Identifying and interpreting variables. In D. R. Preston (ed.), *American Dialect Research* (pp. 193–221). Philadelphia/Amsterdam: John Benjamins.

Wolfram, W. (2003). Reexamining the development of African American English: evidence from isolated communities. *Language*, 79, 282–316.

Wolfram, W. (2006). Variation in language: overview. In K. Brown (chief ed.), *Encyclopedia of Languages and Linguistics*, 2nd edn (pp. 333–40). Oxford: Elsevier.

Wolfram, W. and Christian, D. (1976). *Appalachian Speech*. Washington, DC: Center for Applied Linguistics.

Wolfram, W., Hazen, K., and Schilling-Estes, N. (1999). *Dialect Change and Maintenance on the Outer Banks*. Tuscaloosa, AL: University of Alabama Press.

Wolfram, W. and Schilling, N. (2016). *American English: Dialects and Variation*, 3rd edn. Malden, MA: Wiley Blackwell.

Wolfram, W. and Thomas, E. R. (2002). *The Development of African American English*. Malden, MA: Blackwell.

26 Multilingualism

SUZANNE ROMAINE

1 Introduction

Multilingualism is not the aberration or minority phenomenon supposed by many English speakers. It is, on the contrary, a normal and unremarkable necessity for the majority of the world's population. Despite the spread of global languages, multilingualism remains a reality for most of the world. Around two-thirds of the world's population is bilingual (Baker and Prys Jones 1998: vii).

The existence of roughly 6,900 languages in the world but only about 200 nation-states means there are approximately 30 times as many languages as there are countries and that bilingualism is present in practically every country in the world, whether recognized or not.

It is thus monolingualism which represents a special case, despite the fact that most linguists have paid more attention to it and have taken it to be the norm in their theories of language. As Romaine (1995: 1) remarks, "It would certainly be odd to encounter a book with the title, *Monolingualism*." Nevertheless, the predominantly monolingual orientation of linguistic theory has left its mark on both researchers' terminology and their theoretical concerns. Chomsky (1965: 3), for instance, defined the scope of reference for the study of language as follows: "Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly." By contrast, in the heterogeneous communities where multilinguals reside, individuals are rarely equally fluent in the languages they know. Indeed, a society that produced such individuals would soon cease to be multilingual since no community uses two or more languages for the same set of functions (see Section 4).

In this chapter the terms "bilingualism" and "multilingualism" will be used interchangeably to refer to the use of two or more languages. Because multilingualism exists within the cognitive systems of individuals, as well in as families, communities, and countries, it is perhaps inevitable that the study of various aspects of the phenomenon have been parceled out among various subdisciplines of linguistics and related fields of research such as psychology, sociology, and education, to name just a few. For instance, the acquisition of proficiency in another language usually results in some degree of bilingualism, yet its study is generally regarded as the province of a separate subdiscipline called second language acquisition (see Chapter 27).

Psychologists, for their part, have investigated the effects of bilingualism on mental processes, while sociologists have treated bilingualism as an element in culture conflict and have looked at some of the consequences of linguistic heterogeneity as a societal phenomenon. Educationists have been concerned with bilingualism in connection with public policy. Basic questions about the relationship between bilingualism and intelligence, whether certain types of bilingualism

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

are good or bad, and the circumstances under which they arise, also impinge on education. Within the field of international studies, bilingualism is seen as an essential element in cross-cultural communication. In each of these disciplines, however, multilingualism is too often seen as incidental and has been treated as a special case or as a deviation from the norm.

Within the field of linguistics increasing attention has been given to the systematic study of language contact and the term "contact linguistics" is now used in a wide sense to refer to both the process and outcome of any situation in which two or more languages are in contact. A related field of research has focused on particular types of languages called pidgins and creoles which have emerged in instances where the groups in contact do not learn each other's language or some other language of wider communication already in existence (Romaine 1988). Linguists who study language contact often seek to describe changes at the level of linguistic systems in isolation and abstraction from speakers, thus losing sight of the fact that the bilingual individual is the ultimate locus of contact, as Weinreich (1968) pointed out decades ago. More than half of the nearly 400 million people around the world who speak Spanish, for example, do so in situations of intensive contact with other languages (Silva-Corvalán 1995: 453).

Despite the normality of multilingualism and the centrality of language contact to the evolution of language, research on multilingualism has played a relatively marginal role within linguistics until the last few decades. A variety of textbooks, handbooks, and compendia now offer useful overviews of various aspects of the topic of multilingualism (see, e.g., Auer and Wei 2007; Baker 1996; Baker and Prys Jones 1998, Bhatia and Ritchie 2004; Edwards 1994; Grosjean 1982, 2010; Hakuta 1986; Hoffman 1991; Romaine 1995). Here coverage will be confined to the following topics: origins of multilingualism: causes and consequences; individual vs. societal bilingualism; language choice; language shift and death. The chapter concludes with observations about the changing character of multilingualism in the world today.

2 Origins of Multilingualism: Causes and Consequences

Multilingualism is a condition of life of considerable antiquity, possibly as old as the human species. With the rare exception of small isolated atoll communities, almost none of which are really isolated anymore, human communities were always in contact with other groups and connected to them either economically or socially through exchange of goods, knowledge, marriage partners, etc. Despite the utility of multilingualism in human cultural exchange, the account from *Genesis* 11:1–9 would have us believe that linguistic diversity is the curse of Babel. In a primordial time people spoke the same language. God, however, decided to punish them for their presumptuousness in erecting the tower by making them speak different languages. Thus, multilingualism became an obstacle to further cooperation and placed limits on human worldly achievements. Babel's legacy has given rise to some misinformed views on multilingualism which are still commonplace. It is still widely believed that multilingualism is divisive, and that monolingualism is required for national unity and development.

It takes but little reflection to find flaws in this reasoning and to cite cases where the sharing of a common language has not gone hand in hand with political or indeed any other kind of unity. Northern Ireland is one such example from the English-speaking world which comes readily to mind, but there are many others from other parts of the globe. Certainly, the attempt at Russification of the former republics of the Soviet Union did not ensure unity in that part of the world either. Indeed, one of the first political acts undertaken by the newly independent Baltic states was to reassert their linguistic and cultural autonomy by reinstating their own national languages as official.

Humans have been managing or mismanaging multilingualism for centuries well before modern notions such as "language policy" or "language planning" came onto the scene. Thus, for example, Charles V decided in 1550 to impose Castilian on the Indians of South America.

Long before Europeans came to the island of New Guinea the Motu people of the Papuan coast decided to use a simplified version of their language in their trade contacts with outsiders, and traders in Canton markets wrote numbers on slates to which buyer and seller pointed as they negotiated a price.

What is new, however, are attempts to manage such linguistic and cultural contacts and potential conflicts resulting from them within the framework of agencies of the modern nation-state. As the bedrock of the current political world order, the nation-state is the most critical unit of analysis because it is policies pursued within national boundaries that give some languages (and their speakers) the status of majority and others that of minority (Romaine 2009). What are perceived as problems surrounding multilingualism today are not "natural" problems inherent to multilingualism itself but are largely consequences of a culturally particular construal of the relationship between nations and languages that emerged during the late eighteenth and early nineteenth centuries when new national identities along with the languages and cultures linked to them were created out of the politics of European nation-building. The newly emergent national languages were in effect those of dominant ethnic groups. Today the 28 current members of the European Union recognize 24 languages as official, with most endorsing official monolingualism. Only six (Belgium, Finland, Ireland, Cyprus, Luxemburg, and Malta) are officially bi- or multilingual. Although the mother tongue of the majority of Europeans is one of the state languages of their country, at least 40–50 million citizens of Europe speak a language other than official languages of the state of which they are citizens. The languages spoken by these minorities (not including immigrants) often do not have the same rights as those granted to the official languages. Many indigenous people today like the Bretons and Basque find themselves living in nations that they had no say in creating and are controlled by groups who do not represent their interests and in some cases, actively seek to exterminate them, as was long the case with the Kurds in Iraq and Turkey.

In eighteenth-century France, for example, the possession of a common language was seen as the key to the egalitarian aims of the French Revolution. Speaking French meant being able to participate on equal terms in the newly established French nation-state. The idea of national unity was that France was to become bound together by common goals, administration, and culture. The French language was and still is symbolic of this unity. Since the revolution French nationalists have seen the persistence of non-French speaking groups and their cultures as threats to the stability and persistence of the union.

However, even by 1863 at least one-fifth of the population was still not French-speaking. As late as 1922 the General Inspector of Schools was to declare linguistic war on Bretons who persisted in speaking their own language: "It is of first order importance that Bretons understand and speak the national language: they will only truly be French on that condition . . . It is Frenchmen that are needed to Frenchify the Bretons, they will not Frenchify themselves by themselves" (cited in Kuter 1989: 77). A few years later the Minister of National Education said that "for the linguistic unity of France, the Breton language must disappear" (cited in Kuter 1989: 78). Even today some of the modest attempts to give Breton a limited place in the education system have been resisted by those who feel that any concessions to Bretons will inevitably lead to political separatism. While Mitterand's Socialist government of the 1980s and early 1990s issued a cautious recognition of France as a multicultural nation, it still advocated the fusion of cultures. Ironically, Mitterand saw the resurgence of interest in regional language and culture as an effective force against the increasing influence of American popular culture (see Section 6), while Sarkozy's administration in the early twenty-first century increased demands for assimilation of newer minorities such as Moroccans and Tunisians.

The boundaries of modern nation-states in Africa and other parts of the New World have been arbitrarily drawn, with many of them created by the political and economic interests of Western colonial powers. With the formation of these new nation-states, the question of which language (or which version of a particular one) would become the official language arises and has often

led to bitter controversy. Even countries with more than one official language, such as Canada (where French and English share co-official status), have not escaped attempts by various factions to gain political advantage by exploiting issues of language loyalty.

Some political scientists and linguists have used the term "Fourth World" to label indigenous dispossessed minority peoples who have been encapsulated within, and in some cases divided across, modern nation-states, e.g., the Sami and Inuit peoples of the Arctic region. They are people without their own nation-state, but who nevertheless regard themselves as ethnically and linguistically distinct from the majority population in the countries where they reside. Their struggle for the right to use their own languages continues too. More than 80 percent of the conflicts in the world today are between nation-states and minority peoples (Clay 1990).

Although multilingualism itself is often blamed for these conflicts, language is really a symbol of a much larger struggle for the recognition of minority rights. In 1951 Frisian language activists were involved in a street riot in the Dutch town of Leeuwarden protesting the inadmissibility in Dutch courts of the Frisian language spoken by many of the members of the major indigenous minority group.

3 Individual vs. Societal Multilingualism

Linguists usually draw a distinction between individual and societal multilingualism, although it is not always possible to maintain a strict boundary between the two. Some countries, like Canada, are officially bilingual in English and French, although not all Canadians are bilingual. Many more French-speaking Canadians learn English as a second language than English-speaking Canadians learn French. In other countries such as India, Singapore, and Papua New Guinea high degrees of individual bilingualism mean that the average person knows at least two or more languages. In Singapore four languages, English, Mandarin, Tamil, and Malay share co-official status, and most people are bilingual in English and one of the other official languages.

Some of the connections between individual and societal bilingualism become evident when we consider some of the reasons why certain individuals are or become bilingual. Usually the more powerful groups in any society are able to force their language upon the less powerful. In Britain, the child of English-speaking parents does not have to learn Panjabi or Scottish Gaelic, but speakers of Panjabi and Scottish Gaelic are expected to learn English. In Papua New Guinea few children know English before coming to school, yet most will still be educated in English because this language policy is a legacy of the country's colonial heritage. By contrast, however, the middle-class anglophone parents in Canada who send their child to a French immersion school are under no obligation to do so. Many nevertheless choose such schooling as a means of enriching their children's development because they believe knowledge of another language is an advantage. In Singapore the co-official status granted to Tamil and Malay (also designated the national language) alongside English and Mandarin is not matched by supportive language policies guaranteeing their transmission. School outcomes clearly reflect the advantages being given to the Chinese majority (Gupta 1994).

Even in countries where minority languages are recognized for some purposes, what this means varies in practice. By "minority language" I mean one with a relatively small number of speakers living within the domain of a more widely spoken language, whose knowledge is usually necessary for full participation in society. Swedes in Finland probably have the best legal protection of any minority group in the world. The next strongest position is held by minority languages with limited (often territorial) rights. This is the case in Canada, where certain provinces are officially declared bilingual, and others, like Ontario (where the national capital lies), are not.

It would be naive, however, to assume that bilingual countries were created to promote bilingualism, rather than to guarantee the legal right to more than one language in a society. There is

an important distinction between de facto ("by fact") and de jure ("by law") bilingualism. There are often fewer bilingual individuals in de jure multilingual or bilingual states than in those where de facto multilingualism or bilingualism occurs. A good example is Switzerland, where territorial unilingualism exists under federal multilingualism. Although Switzerland is widely cited as a successful example of multilingualism, only about 6 percent of Swiss citizens can be considered multilingual in the country's four official national languages (German, French, Italian, and Romantsch). English is much preferred over the other official languages as a second language. Of the 26 cantons, 22 are officially monolingual. Economic and political power is more greatly concentrated among German speakers.

4 Language Choice in Multilingual Communities

In all multilingual communities speakers switch among languages or varieties just as monolinguals switch among styles. The fact that speakers select different languages or varieties for use in different situations shows that not all languages/varieties are equal or regarded as equally appropriate or adequate for use in all speech events. A foreigner who manages to learn a variety of Telegu sufficient to get by on the streets of Hyderabad will soon find out that this particular form of Telegu cannot be used for all purposes which an English monolingual might use English for. The average educated person in Hyderabad may use Telegu at home, Sanskrit at the temple, English at the university, Urdu in business, etc. He or she may also know other varieties of Telegu, or Kannada, Tamil, or Malayalam for reading, dealing with servants, or other specific purposes. Many South Asians have active control over what amounts to complex linguistic repertoires drawn from different languages and varieties. In societies such as these, multilingualism is not an incidental feature of language use, but a central organizing force in everyday life. In most parts of India, monolingualism would be problematic relative to the norms and expectations about the number of languages and varieties a person needs in order to manage the everyday things a normal person has to do.

Although language choice is not arbitrary, not all speech communities are organized in the same way. Through the selection of one language over another or one variety of the same language over another speakers display what may be called "acts of identity," choosing the groups with whom they wish to identify (Le Page and Tabouret-Keller 1985). There are, however, some common motivations for such choices in different societies. The first step in understanding these choices is to gain some idea of what languages and varieties are available to speakers in a particular social context. Context in this case may be thought of in its widest sense as the varieties available either officially or unofficially within the boundaries of a nation-state such as Canada, or in a very narrow sense, as those available on a particular occasion, e.g., shopping in an urban market in Kenya, or in a department store in Strasbourg.

4.1 Domains of use

The concepts of domain and diglossia are helpful in understanding the ways in which linguistic resources are organized in multilingual communities, and the factors influencing speakers' choices. In research on the Puerto Rican community in New York City, a team of sociolinguists arrived at a list of five "domains" in which either Spanish or English was used consistently (Fishman, Cooper, and Ma 1971). These were established on the basis of observation and interviews and comprised family, friendship, religion, employment, and education. These domains served as anchor points for distinct value systems embodied in the use of Spanish as opposed to English. A domain is an abstraction referring to a sphere of activity representing a combination of specific times, settings, and role relationships. They conducted further studies to support their claim that each of these domains carried different expectations for using Spanish or English.

They constructed hypothetical conversations that differed in terms of their interlocutors, place, and topic. The way in which these variables were manipulated determined the extent to which the domain configuration was likely to be perceived as congruent or incongruent. For example, a highly congruent configuration would be a conversation with a priest, in church, about how to be a good Christian. A highly incongruent one would be a discussion with one's employer at the beach about how to be a good son or daughter.

People were asked to imagine themselves in hypothetical situations where two of the three components of the conversational context were given. For example, they might be asked to imagine they were talking to someone at their place of work about how to do a job most efficiently. They were then asked to whom they would most likely be talking and in what language. The respondents tended to provide congruent answers for any given domain, and their choice of language was consistent. The most likely place for Spanish was the family domain, followed by friendship, religion, employment, and education.

In each domain there may be pressures of various kinds, e.g., economic, administrative, cultural, political, religious etc., influencing bilinguals towards use of one language over another. Often knowledge and use of one language is an economic necessity. Such is the case for many speakers of a minority language, like Gujerati in Britain, or French in provinces of Canada where Francophones are a minority. The administrative policies of some countries may require civil servants to have knowledge of a second language. For example, in Ireland, knowledge of Irish is required for some positions. In some countries it is expected that educated persons will have knowledge of another language. This is probably true for most European countries, and was even more dramatically so earlier in countries like Russian, where French was the language of polite, cultured individuals. Languages like Greek and Latin have also had great prestige as second languages of the educated. As is the case with accent, the prestige of one language over another is a function of the perceived power of those who speak it. A bilingual may also learn one of the languages for religious reasons. Many Muslim children in Britain receive religious training in Arabic.

Due to competing pressures, it is not possible to predict with absolute certainty which language an individual will use in a particular situation. In trying to account for the choices made by Buang speakers in Papua New Guinea, a country with more than 800 languages, we can take as one example Sankoff's (1980: 36) model of the selections made by speakers in terms of social and situational variables in the speech event, e.g., formality, addressee, etc. Speakers have three languages to choose from: Buang, Yabem (an indigenous language associated with different geographical regions), and Tok Pisin, a variety of pidgin/creole English now widely used in the country. Knowledge of Yabem is largely restricted to those Buang who attended mission schools, where the language of instruction was Yabem, spread by Christian missionaries as a lingua franca.

4.2 Diglossia

The choices made by individuals may become institutionalized at the societal level in communities where bilingualism is widespread. Diglossia is a common form of societal bilingualism in which two varieties or languages are functionally compartmentalized. An example can be taken from Arabic-speaking countries such as Egypt or Iraq in which the language used at home may be a local version of Arabic. The language recognized publicly, however, is modern standard Arabic, which takes many of its normative rules from the classical Arabic of the Koran. The standard language is used for "high" functions such as giving a lecture, reading, writing, or broadcasting, while the home variety is reserved for "low" functions such as interacting with friends at home. High (H) and low (L) varieties differ not only in grammar, phonology, and vocabulary, but also with respect to a number of social characteristics, namely, function, prestige, literary heritage, acquisition, standardization, and stability. L is typically acquired at home as a

mother tongue and continues to be used throughout life. Its main uses are in familial and familiar interactions. H, on the other hand, is learned later through schooling and never at home. H is related to and supported by institutions outside the home. The separate domains in which H and L are acquired immediately provide them with separate institutional support systems.

Diglossic societies are marked not only by this compartmentalization of varieties, but also by restriction of access. Entry to formal institutions such as school and government requires knowledge of H. The extent to which these functions are compartmentalized can be illustrated in the importance attached by community members to using the right variety in the appropriate context. An outsider who learns to speak L and then uses it in a formal speech will be ridiculed. Speakers regard H as superior to L in a number of respects. In some cases H is regarded as the only "real" version of a particular language to the extent that speakers claim they do not speak L. Sometimes the alleged superiority is invoked for religious and/or literary reasons. For example, the fact that classical Arabic is the language of the Qu'rān endows it with special significance. In other cases a long literary tradition backs the H variety, e.g., Sanskrit. There is also a strong tradition of formal grammatical study and standardization associated with H. Table 26.1 shows a typical, though not universal, distribution for high and low varieties in diglossia.

The notion of diglossia has been extended to other communities in which the varieties in diglossic distribution have the status of separate languages, such as Spanish and Guaraní (a local language totally unrelated to Spanish) in Paraguay. Spanish serves here as the high variety used for high functions. It is the official language of government and education, although 90 percent of the population speaks Guaraní, which has the status of national language. Diglossia and bilingualism have been stable there and recent attempts to use Guaraní as a medium of education have met with resistance to extending Guaraní from intimate into public domains. The notion of diglossia is also sometimes expanded to include more than two varieties or languages which participate in such a functional relationship, e.g., in Tunisia, French, Classical Arabic, and Tunisian Arabic are in triglossic distribution, with French and Classical Arabic sharing H functions in relation to Tunisian Arabic, and French occupying the role of H in relation to the other two. The term "polyglossia" has also been used to refer to cases such as Singapore where many varieties coexist in a functional relationship. English, Mandarin, Tamil, and Malay share co-official status, but each of these has local L variants. A child who speaks Hokkien at home may be schooled in

Table 26.1 Some situations for high and low varieties in diglossia (adapted from Ferguson 1959).

	High	Low
Religious service	+	
Instructions to servants, waiters etc.	+	
Personal letter	+	
Speech in parliament, political speech	+	
University lecture		+
Conversation with family, friends, colleagues	+	
News broadcast		+
Radio soap opera		+
Newspaper editorial, news story		+
Comedy		+
Poetry		+
Folk literature		+

Mandarin Chinese at school. English also functions as an H variety to the other three because it has more prestige.

The relationship between individual bilingualism and societal diglossia is not a necessary or causal one. Either phenomenon can occur without the other (Fishman 1967). Both diglossia with and without bilingualism may be relatively stable, long-term arrangements, depending on the circumstances. Consider the Old Order Mennonites and Hasidic Jews in the United States, both of whom maintain stable diglossia with bilingualism. They control their own schools and restrict their use of aspects of the dominant society's culture to economic pursuits, and even these are tightly regulated. For example, the Old Order Mennonites use electricity for pasteurization of milk, as required by law, but they do not have it in their homes for refrigeration or for use with farm machinery. The degree to which the outside world is engaged is justified only to the extent that it contributes to the maintenance of the group. By not accepting or implementing the other culture in its entirety, it is kept in strict complementary distribution with their own. English is specifically excluded from home and religious use and encroaches only in a limited way in economic domains. Religious conservatism has similarly aided Hasidic Jews in maintaining Yiddish as their home language, while, in the secular Jewish community, Yiddish is declining and survives primarily among the few remaining elderly speakers.

Stability, however, is a subjective notion. In some cases indigenous languages can be swamped by intrusive ones over a relatively short period of time. This is what has happened to the Aboriginal languages of Australia and the Celtic languages of the British Isles. For example, the Aboriginal population of Tasmania (ca. 3–4,000) was exterminated within 75 years of contact with Europeans. Some linguists predict that if nothing is done, almost all remaining Aboriginal languages will be extinct within the next hundred years (see Sections 5 and 6). Many bilingual situations do not last for more than three generations. Immigrant languages, for instance, have disappeared as their speakers have adopted the language of the new environment. This is true for many speakers of South Asian languages, like Gujerati and Bengali, in Britain, and speakers of Spanish and Korean in the United States. In cases such as these of bilingualism without diglossia, two languages compete for use in the same domains. When speakers are unable to establish the compartmentalization necessary for survival of the L variety, language shift may be unavoidable (see Section 5).

4.3 Code-switching

Although the existence of bilingualism, diglossia, and code-switching have often been cited as factors leading to language death, in some cases code-switching and diglossia are positive forces in maintaining bilingualism. Swiss German and Faroese may never emerge from diglossia, but are probably in no danger of death. In many communities frequent switching between languages serves important functions. Switching between languages is a communicative option available to bilingual members of a speech community on much the same basis as switching between styles or dialects is an option for monolinguals. Using the term "code" as a neutral one (like "variety") does not commit us to taking a decision as to whether the entities concerned constitute languages or dialects.

An early study conducted by Blom and Gumperz (1972), in a rural Norwegian village called Hemnesberget, introduced the concepts of "metaphorical" and "transactional" (sometimes referred to as "nonsituational" vs. "situational") switching. Transactional switching is the type most commonly discussed as being controlled by components of the speech event like topic and participants as examined in the example from Papua New Guinea in Section 4.1 in the discussion of domains. When Hemnesberget residents step up to the post office counter, greetings and inquiries about family tend to be exchanged in the local dialect, while the business part of the transaction, e.g., buying stamps, is carried out in standard Norwegian. This would be an example of transactional

switching. Metaphorical code-switching, however, concerns the various communicative effects the speaker intends to convey. For example, teachers deliver formal lectures in the official standard form of Norwegian, but shift to regional Norwegian dialect when they want to encourage discussion among the students. While the components of the speech event such as speaker, topic, listener, and setting have not changed, a switch in code alters the tone of the interaction.

A symbolic distinction between "we" and "they" is embodied in the choice of varieties. Generally speaking, the minority code tends to be regarded as the "we," and the majority code as the "they" variety. The "we" variety typically signifies in-group, informal, personalized activities, while the "they" variety marks out-group, more formal relations. In this example from Panjabi/English switching in Britain, Panjabi serves to mark the in-group of Panjabi/English bilinguals and English, the out-group: *esi engrezi sikhi e te why can't they learn*? 'We learn English, so why can't they learn [Asian languages]'. Here the speaker makes the point that Panjabi speakers are expected to learn English, but English people are not required to learn their language. The switch from Panjabi to English emphasizes the boundaries between "them" and "us."

Speakers may switch for a variety of reasons, e.g., to redefine the interaction as appropriate to a different social arena, or to avoid, through continual code-switching, defining the interaction in terms of any social arena. The latter function of avoidance is important because code-switching often serves as a strategy of neutrality or as a means to explore which code is most appropriate and acceptable in a particular situation. In many government offices in Canada, it is customary for bilingual employees to answer the telephone by saying "Bonjour, hello" in order to give the caller the option of choosing either language to continue the conversation.

In some multilingual exchanges the question of code choice is not resolved because the parties involved do not agree on definition of the arena. An example from western Kenya where a brother and sister are conversing in the brother's store shows that where code choice has not been regularized, it must be negotiated on the spot. The siblings are used to conversing on home territory as family members and not as store owner and customer. Because the sister wished to conduct the conversation in her brother's store on the basis of their solidarity as brother and sister in order to get special treatment, she chose their shared mother tongue, Lwidakho. The brother, however, wanted to treat his sister as a customer and therefore used Swahili, which is an ethnically neutral choice in this speech community and the unmarked choice for service encounters of this type. The utterances in Lwidakho are underlined in this exchange. In some ways this conversation is like what happens in Hemnesberget, Norway, except that the sister does not switch to Swahili once the greetings are over, and the brother does not switch back to Lwidakho to accommodate his sister. The sister then goes away without everything she had hoped for (Myers-Scotton 1993: 144–5).

Brother: Good morning, Sister
Sister: Good morning.
Brother: Are you alright?
Sister: Yes, just a little.
Brother: Sister, now today what do you need?
Sister: I want you to give me some salt.
Brother: How much do you need?
Sister: Give me sixty cents worth.
Brother: And what else?
Sister: I would like something else, but I've no money.
Brother: Thank you, sister. Goodbye.
Sister: Thank you. Goodbye.

The preference in market transactions in Jerusalem is for multilingualism, as this example shows, when four women soldiers walk up to look at bracelets outside a jewelry store (Spolsky and Cooper 1991: 108–9):

Shopkeeper: You want bracelets?
Soldier 1: How much?
Shopkeeper: You want this one or this one?
Soldier 2 (in Hebrew): Those aren't pretty.
Soldier 1 (in Arabic): That's not pretty?
Shopkeeper (in Arabic, then Hebrew): Pretty. Like women soldiers.

The shopkeeper addresses the women first in English even though they are Israeli soldiers and obviously native speakers of Hebrew. Because Hebrew has a higher status than Arabic in Israel, for the Arab to use Hebrew would signal a subordinate status. By choosing English, he downplays the nationalist dimensions of Hebrew, and opts for the even higher status associated with English. The first soldier accepts this choice of language, allowing the shopkeeper to continue in this more neutral language. The second soldier introduces Hebrew into the exchange to make a comment to her friend. This may be partly a bargaining ploy because she knows the shopkeeper will understand. The first soldier then switches to Arabic, making clear that she is not an English-speaking tourist or non-Arabic speaking shopper who can be taken advantage of. The shopkeeper replies in Arabic and then Hebrew, establishing his own ability to speak Hebrew and reciprocating the soldier's accommodation to his language.

Accommodation is possible here because all parties have the competence to carry on the activity multilingually. Speakers can exercise a choice only to the extent that they can speak a particular language well enough to choose it over some other in a particular domain. As noted in Section 1, multilinguals rarely develop equal fluency in all the languages they know. Nevertheless, there has been a tendency to regard bilingual competence as the sum of competences in each of the two languages rather than as a unitary system allowing pooling of resources across both. If the proficiency of a bilingual is evaluated in circumstances where s/he is forced to stay within one code, e.g., in contacts with a monolingual community, then that person's communicative competence will seem less rich than it actually is. Greater proficiency in one language in a particular domain may prompt code-switching to that language or result in interference from that language in the less well known language. The repertoires of multilingual speakers can be exploited fully in those multilingual settings where they can draw upon resources from each of the available codes plus strategies for switching between them. Thus, this example from a Malay/English bilingual recorded by Ozog (1987) is totally unremarkable in the everyday life of the speaker concerned, as odd as it may appear to a monolingual. It is difficult to say whether this utterance is basically English with some Malay words, or a Malay utterance with English words.

This morning I hantar my baby tu dekat babysitter tu lah.
'This morning I took my baby to the babysitter.'

Popular attitudes towards some kinds of code-switching, mixing, and interference are, nevertheless, often negative, even among community members themselves who engage in this kind of multilingual behavior frequently. Indeed, in the Panjabi-speaking community in Britain many people label examples of the type cited above as *tuti-futi* ("broken up") Panjabi and do not consider it to be "real" Panjabi (Chana and Romaine 1984). In parts of French-speaking Canada the term "joual" has similar connotations. In communities like these there is almost an inherent conflict between the desire to adopt English loanwords as prestige markers and their condemnation as foreign elements destroying the purity of the borrowing language. Haugen (1977: 332) described the ambiguity felt by Norwegian Americans who did not approve of people from their own group who tried to speak too bookishly, but at the same time they poked fun at those who adopted excessive numbers of English words, calling them "yankeefied." A visitor from Norway commented on hearing this American variety of Norwegian that it was "no language whatever, but a gruesome mixture of Norwegian and English, and often one does not know whether to take it humorously or seriously" (Haugen 1977: 94).

A change in political consciousness, however, may lead to a change in attitudes with the result that code-switching is taken very seriously. In parts of the southwestern USA and California, where code-switching between Spanish and English is frequent among the Hispanic population, terms such as "Tex-Mex," "español tuerco" ("twisted Spanish"), or "mocho" ("broken") and "Spanglish" are used to refer to mixed varieties of Spanish/English. While the terms still have derogatory overtones in some quarters, these mixed codes have come to serve as positive ethnic markers of Hispanic identity and are increasingly used in literature. Gloria Anzaldúa (1999: 77), for example, wrote about how she found herself at odds with the community of both English and Spanish speakers who regarded her language as wrong, but that, until she could accept the legitimacy of Tex-Mex, she could not accept her own legitimacy.

5 Language Shift and Death

Choices made by individuals on an everyday basis can also have an effect on the long term relationships of languages in contact. Language shift generally involves bilingualism (often with diglossia) as a stage on the way to eventual monolingualism in a new language. Typically a once monolingual community becomes bilingual as a result of contact with another (usually socially more powerful) group and becomes transitionally bilingual in the new language until their own language is given up altogether. A comparative examination of 35 nations revealed that in no other country was the rate of shift toward monolingualism in the national dominant language as fast as in the United States (*Lieberson, Dalto, and Johnston 1975*). This shift is equally as fast as it was in past generations (Romaine 2010a; Veltman 2000). No immigrant group, with the possible exceptions of isolated communities like the above-mentioned Old Order Mennonites or Hasidic Jews, has been able to preserve its minority language longer than two or three generations. For example, among German-speaking Anabaptist immigrants such as the Amish, Hutterites, and Mennonites, degree of religious conservatism corresponds with extent of retention of German. Among less conservative groups, shift to English is swift and complete as soon as English intrudes into what were German domains.

In Australia the decline of non-English languages has been similarly dramatic. Only 4.2 percent of the Australian-born population regularly uses a language other than English. This figure includes Aboriginal languages too. Yet there are some major differences in the extent to which native languages are retained by the different ethnic groups. Greek Australians display the greatest maintenance, and Dutch Australians the least. Different languages are concentrated in different states, although there is no single minority language of equal significance to Spanish in the USA (Romaine 1991). The United States is now the fifth largest Spanish-speaking country in the world (see further in Section 6).

There are many reasons for language shift and death. In some cases shift occurs as a result of forced or voluntary immigration to a place where it is not possible to maintain one's native language, e.g., Italians in the United States, or as a result of conquest, e.g., the Gaels in Scotland and Ireland. The ultimate loss of a language is termed "language death." Among the many factors responsible for language shift and death are religious and educational background, settlement patterns, ties with the homeland (in the case of immigrant bilingualism), extent of exogamous marriage, attitudes of majority and minority language groups, government policies concerning language and education, etc. While each of these factors may be implicated in shift and death, they do not entirely determine the fate of a language (Romaine 2010b).

Where large groups of immigrants concentrate in particular geographical areas, they are often better able to preserve their languages, e.g., third generation Chinese Americans residing in Chinatowns have shifted less towards English than their age mates outside Chinatowns. Often a shift from rural to urban areas triggers a language shift. For example, in Papua New Guinea, where Tok Pisin is the language most used in the towns, many children grow up not speaking

their parents' vernacular languages. When a language serves important religious functions, as German does among Old Order Mennonites, it may stand a better chance of survival.

The inability of minorities to maintain the home as an intact domain for the use of their language has often been decisive for language shift. There is a high rate of loss in mixed marriages, e.g., in Wales, where, if Welsh is not the language of the home, the onus for transmission is shifted to the school. Identification with a language and positive attitudes towards it cannot guarantee its maintenance. In Ireland the necessity of using English has overpowered antipathy towards English and English speakers. In some cases speakers may be forbidden to use their language altogether, as Kurds in Turkey were until recently. In a community whose language is under threat, it is difficult for children to acquire the language fully. Languages undergoing shift often display characteristic types of changes such as simplification of complex grammatical structures. These changes are the result of decreased use of the language in certain contexts, which may lead to a loss of stylistic options. In some Native American Indian languages of the southwestern United States complex syntactic structures have become less frequent because the formal and poetic styles of language are no longer used. The degree of linguistic assimilation may serve as an index of social assimilation of a group. It depends on many factors such as receptiveness of the group to the other culture and language, possibility of acceptance by the dominant group, degree of similarity between the two groups, etc. Albanian speakers who emigrated to Greece have more readily given up their language and assimilated than have Albanian speakers in Italy, where attitudes towards diversity are more favorable.

There is no doubt that absence of schooling in one's own language can make maintenance difficult. In a study done of 46 linguistic minorities in 14 European countries, the clearest link to emerge between language and schooling is that a minority language which is not taught tends to decline (Allardt 1979). Studies of language shift have shown repeatedly that schools are a major agent of cultural and linguistic assimilation. Formal education is often the first point of contact children have with the world outside their own community. English schools were destructive to Dyirbal-speaking children in Aboriginal Australia for several reasons. The very fact that Dyirbal had no presence in the school was a signal that it was seen as a useless language. Schools also provide a major context for the use of English and exposure to English-speaking children. By being immersed into a totally English environment, the Dyirbal child was denied the opportunity of learning in Dyirbal. An educational program of this type is called "submersion" because the child's native language is suppressed or ignored and the children have to sink or swim in a completely different language environment. The aim of such programs is cultural and linguistic assimilation.

In many parts of the world today children are not taught enough of their own language and culture to appreciate it. They become caught in a vicious circle. Because the school fails to support the home language, skills in it are often poor. The failure of the school to let children develop further in their own language is then used to legitimize further oppression of it. At the same time they do not progress in the majority language (often for reasons that have nothing to do with language, but which reflect the poorer socioeconomic status of the minority in relation to the majority). The economic returns from schooling are greater for those who are advantaged to begin with.

Older Saami in Finland were indoctrinated by the school system into believing that speaking Saami even at home weakened their children's knowledge of Finnish (Aikio 1984). Uninformed school and healthcare officials continue to dispense such advice and are also likely to condemn language mixing and code-switching as harmful to children's development. Beliefs about bilingualism causing stuttering and delayed onset of language are also widespread, despite lack of evidence for them. Most of the so-called experts who offer such advice are monolinguals and think of bilingualism as a problem in need of remediation because they are uninformed about the course of normal bilingual development. National education policies are radically out of line with the realities of multilingualism (UNESCO 2003). The majority of countries in the world

actually operate either de facto or de jure as monolingual in recognizing only one language for use in education; only a quarter recognize more than one language (Edwards 2007: 44). In showing a preference for some language(s), whether designated as official or national or not, the state's decision benefits those for whom the chosen language(s) is a primary language.

A substantial body of research has established the value of high levels of bilingualism for cognitive flexibility and the advantages of teaching children through their mother tongue for a long as possible (Bialystock 2007; Dutcher and Tucker 1997). Results from so-called immersion programs aimed at enriching children's skills in their own language have shown positive benefits in many parts of the world. In Hawai'i, for instance, after nearly a hundred years of decline of the Hawaiian language among the native Hawaiian-speaking population, a program of Hawaiian immersion preschools was set up in 1984 called Pūnana Leo ("language nest"), modeled on a similar program for Maori immersion in New Zealand. This was the first indigenous language immersion program in the USA. At that time there were fewer than 1,000 Hawaiians in the state speaking the language. Fewer than 30 of those were under 18. The program has since been expanded to elementary education in 1987, to secondary education in 1995, and to university level in 1999.

However, school is only one and probably not the most important of all the societal institutions that contribute to and are responsible for language acquisition and maintenance. It would be wrong to leave the picture unduly rosy. Provision of schooling in a minority language will not automatically safeguard its future. Nowhere have language movements succeeded if they expected the school or state to carry the primary burden of maintenance or revival. It is no accident that minority groups who have retained control over their schooling such as the Old Order Mennonites in Pennsylvania have shown greater language maintenance than those who have not.

6 The Changing Face of Multilingualism in the Modern World

At least three trends deserve comment in relation to the changing character of multilingualism in the world today. One is increasing bilingualism in a metropolitan language, particularly English, as numerous smaller languages are dying out due to the spread of a few world languages such as English, French, Spanish, Chinese etc. (Nettle and Romaine 2000). Only nine languages have more than 100 million speakers; altogether they are spoken by about 50 percent of the world's population (Lewis 2009). In this respect, the majority of the world's languages are minority languages. The second is increasing linguistic and cultural diversity the world over through continuing and new waves of immigration.

These two processes represent a struggle between increasing internationalization, cultural and linguistic homogenization (Coca-colonization, as it has sometimes been referred to) vs. diversification. There is a clash of values inherent in the struggle between the global and local, between uniformity and diversity. The language of McWorld is English: Not to use it is to risk ostracization from the benefits of the global economy. It is for this reason that many developing countries opted to use the language of their former colonizers rather than try to develop their own language(s). Using English or French in Africa seems to be cheaper than multilingualism. Such utilitarian methods of accounting do not, of course, factor in the social cost of effectively disenfranchising the majority of citizens who do not know English or French in many Third World nations where these are the official languages. Some 90 percent of Africans have no knowledge of the official language of their country and Africa has the highest proportion of people (87%) without access to mother tongue education (Mackey 1989: 5). Policies privileging international over local languages lead to cultural poverty when linguistic diversity is lost. When large portions of the population are denied forms of self-expression, the nation's political and social

foundations are weakened. A nation that incorporates cultural and linguistic diversity is also richer than one which denies their existence. It can easily be shown that denying people the right to their own language and culture does not provide a workable solution either.

Ethnicity often grows stronger when actively denied, oppressed, or repressed. Throughout its 74 years of existence the former Yugoslavia was a powder keg of ethnic rivalries going back centuries, and held in check by a centralized Communist government until 1980. Once the old regime crumbled, old tensions could surface, leading to the unraveling of the country. The virtual collapse of the economies of the former Soviet bloc countries has shown the difficulties of centralized planning which rides rough shod over the regional and ethnic affiliations.

As far as the trend towards increasing diversity as a result of new immigration is concerned, in the European Union, for instance, ethnic minorities already comprise one third of the urban population under the age of 35 and 10 percent of the school age population has a culture and language different from that of the majority of the country in which they reside. This figure obscures wide variation among member states. In the Netherlands the influx of ethnic minority children in elementary schools in the four largest cities increased to more than 50 percent in the year 2000 (Extra and Gorter 2007). In the USA projections for increasing diversity in the twenty-first century indicate that Hispanics alone may comprise over 30 percent of the total population. If we calculate the long-range social and economic cost of continuing the present pattern of undereducating minority children in Europe and the USA, the results are enormous. It is these children who will become the majority and upon whom the economic burden will fall of caring for the next generation of children and the previous generation, soon to be retirees. At the same time the highly developed technological economies in Europe and the USA will require an increasingly highly educated workforce. New member states in the European Union are almost certain to bring with them their own unresolved language problems and tensions between majority and minorities. Thus, conflicts will likely increase rather than decrease.

In most parts of the world, however, there is still little enthusiasm for the languages of immigrant minorities, even when the language concerned is a world language such as Spanish (as is the case in the USA) or Arabic (as is the case in France and the Netherlands). This is due to status differences between majority and minority populations. Distinctive food, dress, song, etc. are often accepted and allowed to be part of the mainstream, but language seldom is. Another irony in the resistance to providing support in the form of home language instruction to immigrant pupils is that opposition to it in the USA has occurred side by side with increasing concern over the lack of competence in foreign languages. Thus, while foreign language instruction in the world's major languages in mainstream schools has been seen as valuable, both economically and culturally, bilingual education for minority students has been equated with poverty, and loyalties to nonmainstream culture threatening the cohesiveness of the state. Writing about the supposed challenges to American national identity, Huntington (2004: 256) contended that "There is no Americano dream. There is only the American dream created by an Anglo-Protestant society, and Mexican-Americans will share in that dream and in that society only if they dream in English."

7 Conclusions

Monolingual thinking in a de facto multilingual world has rendered linguistic diversity problematic, as suggested by Baetens-Beardsmore (2003: 10) when he observed that "there is a deep-seated and widespread fear of bilingualism. Moreover, there is an all-pervading tendency to couple the notion of "problems" to that of bilingualism, a connotation that never comes to mind in discussions on unilingualism." Nevertheless, there is no evidence indicating that multilingualism is an inherently problematic mode of organization, either for a society or for an individual. Because languages and dialects are often potent symbols of class, gender, ethnic, and

other kinds of differentiation, it is easy to think that language underlies conflict in multilingual societies. Disputes involving language are really not about language, but instead about fundamental inequalities between groups who happen to speak different languages. Multilingualism is shaped in different ways depending on a variety of social and other factors which must be taken into account when trying to assess the skills of speakers and how speakers use the languages they know. It is possible for a bilingual to be fluent in both languages taken together without being able to function completely like a monolingual in either one on its own. The study of the behavior of multilingual individuals and societies thus requires us to go beyond many of the concepts and analytical techniques presently used within linguistic theory designed for the description of monolinguals.

REFERENCES

Aikio, M. (1984). The position and use of the Same language: historical, contemporary and future perspectives. *Journal of Multilingual and Multicultural Development*, 5, 277–91.

Allardt, E. (1979). *Implications of the Ethnic Revival in Modern Industrialized Society. A Comparative Study of the Linguistic Minorities in Western Europe*. Helsinki: Societas Scientariarum Fennica.

Anzaldúa, G. (1999). *Borderlands/La Frontera: The New Mestiza*, 2nd edn. San Francisco, CA: Aunt Lute Books.

Auer, P. and Wei, L. (eds) (2007). *Handbook of Multilingualism and Multilingual Communication*. Berlin: Mouton de Gruyter.

Baetens-Beardsmore, H. (2003). Who's afraid of bilingualism? In J.-M. Dewaele, L. Wei, and A. Housen (eds), *Bilingualism: Beyond Basic Principles* (pp. 10–27). Clevedon: Multilingual Matters.

Baker, C. (1996). *Foundations of Bilingual Education and Bilingualism*, 2nd edn. Clevedon: Multilingual Matters.

Baker, C. and Prys Jones, S. (1998). *Encyclopedia of Bilingualism and Bilingual Education*. Clevedon: Multilingual Matters.

Bhatia, T. K. and Ritchie, W. C. (eds) (2004). *The Handbook of Bilingualism*. Oxford: Blackwell.

Bialystock, E. (2007). Cognitive effects of bilingualism: how linguistic experience leads to cognitive change. *The International Journal of Bilingual Education and Bilingualism*, 10, 3, 210–23.

Blom, J. P. and Gumperz, J. J. (1972). Social meaning in linguistic structures: code-switching in Norway. In J. J. Gumperz and D. Hymes (eds) *Directions in Sociolinguistics* (pp. 407–34). New York: Holt, Rinehart and Winston.

Chana, U. and Romaine, S. (1984). Evaluative reactions to Panjabi/English code-switching. *Journal of Multilingual and Multicultural Development*, 6, 447–73.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Clay, J. W. (1990). Indigenous peoples: the miner's canary for the twentieth century. In S. Head and R. Heinzman (eds) *Lessons of the Rainforest* (pp. 106–17). San Francisco, CA: Sierra Club Books.

Dutcher, N. and Tucker, G. R. (1997). *The Use of First and Second Languages in Education: A Review of Educational Experience*. Washington, DC: World Bank.

Edwards, J. R. (1994). *Multilingualism*. London: Routledge.

Edwards, J. R. (2007). Societal multilingualism: reality, recognition and response. In P. Auer and L. Wei (eds), *Handbook of Multilingualism and Multilingual Communication* (pp. 447–69). Berlin: Mouton de Gruyter.

Extra, G. and Gorter, D. (2007). Regional and immigrant minority languages in Europe. In M. Hellinger and A. Pauwels (eds), *Language and Communication: Diversity and Change* (pp. 15–52). Berlin: Mouton de Gruyter.

Ferguson, C. A. (1959). Diglossia. *Word*, 15, 325–40.

Fishman, J. A. (1967). Bilingualism with and without diglossia; diglossia with and without bilingualism. *Journal of Social Issues*, 23, 29–38.

Fishman, J. A., Cooper, R. L., and Ma, R. (1971). *Bilingualism in the Barrio*. Bloomington, IN: Indiana University Press.

Grosjean, F. (1982). *Life with Two Languages. An Introduction to Bilingualism*. Cambridge, MA: Harvard University Press.

Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press.

Gupta, A. F. (1994). *The Step-Tongue: Children's English in Singapore*. Clevedon: Multilingual Matters.

Hakuta, K. (1986). *Mirror of Language. The Debate on Bilingualism*. New York: Basic Books.

Haugen, E. (1977). Norm and deviation in bilingual communities. In P. A. Hornby (ed.), *Bilingualism:*

Psychological, Social and Educational Implications (pp. 91–103). New York: Academic Press.
Hoffman, C. (1991). *An Introduction to Bilingualism*. London: Longman.
Huntington, S. (2004). *Who Are We? The Challenges to American National Identity*. New York: Simon & Schuster.
Kuter, L. (1989). Breton v. French: language and the opposition of political, economic, social, and cultural values. In N. C. Dorian (ed.), *Investigating Obsolescence: Studies in Language Contraction and Death* (pp. 75–89). Cambridge: Cambridge University Press.
Le Page, R. B. and Tabouret-Keller, A. (1985). *Acts of Identity. Creole-Based Approaches to Language and Ethnicity*. Cambridge: Cambridge University Press.
Lieberson, S., Dalto, G., and Johnston, M. E. (1975). The course of mother-tongue diversity in nations. *American Journal of Sociology,* 81, 34–61.
Lewis, M. P. (2009). *Ethnologue: Languages of the World*, 16th edn. Dallas, TX: SIL International. Also available online at http://www.ethnologue.com/
Mackey, W. F. (1989). Status of languages in multilingual societies. In U. Ammon (ed.), *Status and Function of Languages and Language Varieties* (pp. 3–20). Berlin: Mouton de Gruyter.
Myers-Scotton, C. (1992). *Motivations for Code-Switching: Evidence from Africa*. Oxford: Oxford University Press.
Nettle, D. and Romaine, S. (2000). *Vanishing Voices. The Extinction of the World's Languages*. New York: Oxford University Press.
Ozog, A. C. K. (1987). The syntax of the mixed language of Malay. *RELC Journal*, 18, 72–90.
Romaine, S. (1988). *Pidgin and Creole Languages*. London: Longman.
Romaine, S. (ed.) (1991). *Language in Australia*. Cambridge: Cambridge University Press.
Romaine, S. (1995). *Bilingualism*, 2nd edn. Oxford: Blackwell.
Romaine, S. (2009). Language, culture and identity issues across nations. In J. Banks (ed.), *Routledge International Companion to Multicultural Education* (pp. 373–84). London: Routledge.
Romaine, S. (2010a). Language contact in the USA. In K. Potowski (ed.), *Exploring Linguistic Diversity in the United States* (pp. 25–46). Cambridge: Cambridge University Press.
Romaine, S. (2010b). Language contact and death. In R. Hickey (ed.), *Handbook of Language Contact* (pp. 320–39). Cambridge: Cambridge University Press.
Sankoff, G. (1980). Language use in multilingual societies: some alternate approaches. In G. Sankoff, *The Social Life of Language* (pp. 29–46). Philadelphia, PA: University of Pennsylvania Press.
Silva-Corvalán, C. (ed.) (1995). *Spanish in Four Continents: Studies in Language Contact and Bilingualism*. Washington, DC: Georgetown University Press.
Spolsky, B. and Cooper, R. L. (1991). *The Languages of Jerusalem*. Oxford: Oxford University Press.
UNESCO (2003). *Education in a Multilingual World*. Paris: UNESCO.
Veltman, C. (2000). The American linguistic mosaic: understanding language shift in the United States. In S. L. McKay and S.-L. C. Wong (eds.), *New Immigrants in the United States* (pp. 58–93). Cambridge: Cambridge University Press.
Weinreich, U. (1968). *Languages in Contact*. The Hague: Mouton. (First published 1953. New York: Linguistic Circle of New York Publication No. 2.)

27 Second Language Acquisition

One Person with Two Languages

VIVIAN COOK

Most linguistics has concerned itself with people who know, use or learn a single language. Yet such monolinguals are probably in a minority in the world as a whole. Many people use several languages in the course of a day, whether in multilingual countries such as Pakistan or the Cameroon or in apparently monolingual countries such as England and Japan. This chapter looks at some of the questions raised by the fact that people know more than one language.

Any issue in linguistics can potentially be studied in the context of people who know more than one language; virtually any chapter in this book could then have a parallel form dealing with second languages. The aim is to introduce the reader to some of the issues that have been investigated, touching on areas of linguistics such as phonology and vocabulary as well as syntax. Though a comparative newcomer, second language acquisition (SLA) research is a vast and expanding discipline with its own annual conferences such as EUROSLA (European Second Language Association). A survey by Ellis (2008) is over 800 pages long despite barely touching on areas such as vocabulary or phonology. The account here represents one person's route through a large maze, trying not to stray down paths less connected with linguistics. It starts with general issues in SLA research, it provides some historical background, and then discusses nine interrelated questions that have driven SLA research within the overall context of one person with two or more languages.

1 Overall Issues

The approach here is to look at areas where some relationship can be found between linguistics and SLA research. Much current SLA research is, however, not linguistics-friendly. It often adopts the kind of model of language that linguists have loved to scorn, coming from the tradition of Bloomfieldian structuralist behaviorism and leading up to the new psychology recounted in books such as Tomasello (1998), and embodied in theories, such as chaos theory, emergentism, and connectionism, that do not deal with language in the usual senses recognised by linguists. This rejection of linguistics also largely includes a rejection of the descriptive tools provided by current linguistics. Syntax mostly comes from traditional or descriptive grammar; phonology consists only of the concept of the phoneme; morphology goes no further than the issue of

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

inflections for English verbs and nouns. The baby of useful linguistic tools has been thrown out along with the bathwater of linguistic theories.

Any area that draws on linguistics has to take into consideration the extent to which it is independent of linguistics. Broadly speaking, SLA can be seen either as a testing ground for ideas from linguistics itself or as an independent discipline that draws upon linguistics. People have used second language acquisition as another source of data for trying out linguistic theories and syntactic models, as has been done for first language acquisition, creoles, sign language, and so on. No sooner is a syntactic hypothesis proposed than it gets tested on second language learners. The aim is not so much to answer a research question about second language acquisition as to test out the theory or the overall acquisition model. For example, the accessibility hierarchy for relative clauses (Keenan and Comrie 1977) was tested out on L2 users by Cook (1975) and others. The advocates of this view in a sense see themselves as theoretical linguists who happen to use second language data. The goal is to feed ideas back into mainstream linguistics and influence linguistic theory.

The straight linguist looking at SLA research soon finds that it is using linguistics from every possible school and every possible period. The difficulty is synchronically whether different theories can be combined and diachronically whether different stages of the same theory can be reconciled. Can you combine claims based on Chomsky's Minimalist Program of the 2000s (e.g., Chomsky 1995) with syntactic devices based on the government/binding theory of the 1980s (e.g., Chomsky 1986), as is done everywhere in the generative study of second language acquisition? SLA research tends to treat the ideas it borrows from linguistics, or indeed psychology, as equal and timeless. The phoneme lives on in SLA research long after its death sentence was pronounced by Chomsky and Halle (1968). In addition SLA research often falls back on nonlinguistic traditions of syntactic analysis taken from the English as a foreign language (EFL) grammar tradition of say Palmer (1926) and Hornby (1956), as found in most EFL coursebooks, or from the descriptive grammar tradition of Zandvoort (1957) and Quirk et al. (1972).

Obviously SLA research is perfectly entitled to take what it can from outside and to use it for its own purposes. The fact that cordless drills were first designed to be used on the moon did not mean they could not be used for putting up bookshelves. The use of the phoneme or the accessibility hierarchy is justified by whether they answer questions about second language acquisition, not by whether they form part of linguistics itself. The L2 research is not necessarily invalidated if it does not take on the original premise on which the research technique is founded. Nevertheless there are dangers.

One is when linguistics has moved on from the original idea to something else. Perhaps whether the theory is still current in linguistics is irrelevant to SLA research so long as it suits the needs of SLA research. But linguistics may have abandoned a theory for good reasons. Take the example of the original *Syntactic Structures* model of kernel sentences and transforms (Chomsky 1957; Cook 2007). This was greeted with eagerness by second language acquisition researchers of the period, such as Roulet (1973), and even led to language teaching which forbade the students from using transforms for the first year of teaching. But it would be dangerous for a researcher to employ it today as the model was rapidly succeeded by Aspects, GB, and all their descendants (Cook and Newson 2007). A case does need to be made to justify SLA research using theories or descriptive devices that have been superseded in linguistics on the grounds that the reason for their obsolescence is nothing to do with second language acquisition.

Much of the use of linguistics in SLA research is then paying lip service by borrowing a few general linguistic terms. In a review of the *Handbook of Bilingualism* (Kroll and De Groot 2005), Cook (2008) insisted "Much of the research utilizes descriptive terms such as phonemes ('the building blocks', 72), syllables, words, and meanings, as if these were primitive concepts that were unproblematic and unrelated to any linguistic theory." The terms give the illusion of being linguistics but are more based on popular usage. Since they are rarely defined relative to any linguistic theory and are seldom supported by linguistic descriptions, such old warhorses of SLA research as the grammatical morpheme have rather little connection with linguistics.

The aim of SLA research is then to study people who know two languages. This is arguably not the same as saying that it studies the second language alone. People who know two languages differ in many ways from those who only know one (Cook 2003a). The research questions differ from those involved in first language acquisition since the second language forms part of a complex system within a mind that already has one language; they cannot be phrased solely in terms of a second language without losing the whole rationale for studying second language acquisition – it is *not* first language acquisition by another name.

2 Early Days: Links and Questions

A relationship between linguistics and SLA research started to emerge with the influential distinction made by Weinreich (1953) between compound and coordinate bilinguals. A coordinate bilingual has two separate concepts for a word: two concepts, two words. The concept of "book," for example, can be shown as 🕮, expressed by "book" in the English of English/French bilinguals, or as 🕮, expressed by "livre" in French. The two languages are separate in the mind; a coordinate bilingual may be unable to translate from one language to the other. Compound bilinguals on the other hand have a single concept 🕮 expressed as two different words "book" and "livre": the two languages are tied together in their minds via a common concept: one concept, two words. Though individuals were once thought to be either coordinate or compound bilinguals, more recently it is believed that both types of bilingualism are present to varying extents in the same person; that is say, in a given individual, some aspects of the two languages may be linked, others may be separate. Cook (2003a) proposed a continuum between total integration of the two languages and total separation. This may distinguish one individual from another, may apply differently for different components of language, or may reflect a learner's progress in one direction along the continuum or the other.

The linguistic and behaviorist theories of learning of the 1950s also contributed the concept of transfer to SLA research. Transfer means carrying over the forms and meanings of one language to another, resulting in interference – "instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language" (Weinreich 1953: 1). French users of English fail to distinguish /i:/ and /i/ in "keen" /ki:n/ and "kin" /kin/ because the distinction does not exist in French; Japanese learners of English produce spellings such as **adovocated*, **courese*, and **Engilish*, because there are no consonant clusters in Japanese. The concept of transfer led to the approach called contrastive analysis, which looked for differences between the two languages; these form the main areas of difficulty for learners and automatically lead to "negative" transfer from the L1 (Lado 1957).

Though transfer remains an indispensable concept in SLA research, the contrastive analysis approach itself has mostly been subsumed within other traditions. One reason was that it was all too easy to carry out large-scale comparisons of languages to predict what learners might do, only to find the predicted errors do not occur but other errors do. It was more economical to start from the errors in actual L2 learners' speech and then to work back to their causes – a methodology that became known as error analysis (Corder 1971; James 1998). Bulgarian learners, for instance, produce sentences such as **The my car broke down* (Toncheva 1988); the probable cause is that possessives and articles can occur together in Bulgarian, i.e., their L1 transfer is established by the post hoc error analysis method rather than predicted in advance by contrastive analysis.

The overall issue emerging from these beginnings is how multiple languages relate to each other inside the mind of one person, both during the process of acquiring the L2 and while actually using it. This is reflected in the divergence of definitions for bilingualism itself. At one extreme are "maximal" definitions such as "native-like control of two languages" (Bloomfield 1933), renamed more transparently as "ambilingualism" by Halliday, McIntosh, and Strevens (1964): Bilinguals have as extensive control over their second language as over their first. At the

other extreme are "minimal" definitions that bilingualism starts at "the point where a speaker can first produce complete meaningful utterances in the other language" (Haugen 1953): Any real use of a second language counts as bilingualism, however minimal it may be. Hardly anybody meets the maximal definition since no one commands all the uses of both languages equally well; virtually everybody meets the second definition in that they are capable of using isolated L2 expressions such as "Goodbye," "Bonjour," or "Konichiwa" appropriately. This does however potentially confine the bilingual within the confines of double monolingualism: There are uses of language by bilinguals such as code-switching that are not used by monolinguals of either language.

A strong early influence on SLA research came from the overall structuralist and behaviorist paradigm of linguistics laid down by Bloomfield (1933). When mainstream linguistics swung away from structuralist models, SLA research took on board certain Chomskyan tenets. The key concept was that children have independent grammars of their own that are not simply debased versions of the adult grammar. English L1 children who say **Him go shop* are not reproducing something that they have heard from an adult but have invented a rule of their own that "him" can be a subject. This led to the realization that L2 learners also build up grammars of their own that are not part of either the first language or the second but have an independent existence. L2 learners may say **I not get away* even though putting the negative "not" in front of the verb is found in neither their L2 (English) nor their L1 (German): They too have made up a rule. Several terms were coined for the idea that L2 learners had independent grammars, each with slightly different emphases, such as "approximative system" (Nemser 1971) and "transitional competence" (Corder 1967). The term that found favour was "interlanguage," derived from an eponymous paper by Selinker (1972). When an interlanguage gets stuck at a position where it does not have the characteristics of the target language, it is said to be "fossilized" (Han 2004).

The overall contributions of these early days of SLA research were that the two languages may be separate or closely linked in the mind, that the forms of one language may affect the other, and that L2 learners create a distinct interlanguage with its own rules and properties.

3 What Is the Sequence of L2 Acquisition?

The question that interested many of the first SLA researchers was the order in which people acquire an L2: Is there a "natural" sequence through which all L2 learners progress or does it vary from one person to another, say, according to their L1? One research method was to score the presence of certain English "grammatical morphemes" such as continuous 'ing' going and plural 's' books in the speech of L2 learners. Spanish-speaking children learning English start with plural 's' and progress through continuous 'ing', copula 'be is', and so on (Dulay and Burt 1973). Fairly similar orders were found regardless of whether the learners were in a classroom (Lightbown 1987), whether they were in a country where the language was spoken or not (Makino 1993), and many other factors. Clearly L2 learners of English acquire these grammatical morphemes in a sequence of some kind. However, difficulties emerged with the methodology (was this really the order in which they learned them or simply the order of difficulty?) and with the grammar (do these grammatical morphemes such as verbs 'be', inflections 's', and prepositions 'to' really form a coherent syntactic group?).

More convincing sequences of L2 acquisition emerged from the ZISA project (Meisel, Clahsen, and Pienemann 1981). This studied the development of L2 German by migrant workers with various L1s over a period of time. It was mostly concerned with German word order, which differs from English in that the verb normally comes second in the sentence. Sometimes this yields a Subject Verb Object order, as in *Ich liebe dich* (I love you), but often it leads to an Adverb Verb Subject order, *Immer liebe ich dich* (Always love I you), and to other combinations in which the verb comes second. L2 learners of German start by putting subjects in front of verbs as in *Die*

Kinder essen Apfel (the children eat apple) before they learn how to get other verb-second orders by moving elements around in the sentence, for example Adverb Verb *Da Kinder spielen* (there children play). The sequence of acquisition depends on first acquiring a typical word order and then learning how to move elements about. The essentials of this sequence were duplicated with L2 learners of English in Australia by Pienemann and Johnston (1987): The stage of Subject-Verb-Object *I like apples* preceded the sentences with movement **Apples I like very much*. Learners differ from each other according how much they simplify sentences, say, omitting items such as the copula verb **Ich Mädchen* (I girl) even if they are at the same developmental stage. This "variation" dimension complements the "development" dimension, leading to the "multidimensional model," later known as processability theory (Pienemann 1998).

The concept of sequence was taken further within the large-scale European Science Foundation (ESF) project, which looked at the learning of five L2s by young adult speakers of six L1s (Klein and Perdue 1992, 1997). It found that learners start with sentences without verbs, go on to use verbs without inflection **Its pinch some bread*, and finally have sentences with verbs with inflections **Man is coming in*. Regardless of which language they are learning, the learners arrive at a common basic L2 grammar. A sentence may consist of:

(A) a Noun Phrase followed by a verb followed by an optional Noun Phrase *Mädchen nehme Brot* (girl take bread);

(B) a Noun Phrase followed by a Copula verb followed by a Noun Phrase, Adjective, or Prepositional Phrase *it's bread*;

(C) a Verb followed by a Noun Phrase **pinching its*.

Progress beyond this basic grammar consists largely of fleshing out the different verbal forms. The stages of acquisition derive from communication and pragmatic principles: L2 learners "organise their utterances and texts according to elementary principles of their human language capacity" (Klein and Perdue 1997: 343).

The ESF project is thus a practical demonstration of the interlanguage hypothesis since it shows a common interlanguage independent of both L1 and L2. The project's aim was indeed to see "whether a learner variety is based on recognisable organisational principles, how these principles interact, and whether they also apply to fully-fledged languages" (Klein and Perdue 1992: 1).

Stages of development are not interesting unless they lead to insights into learning. The study of sequences of acquisition thus shades into explanation. For example, Wieden and Nemser (1991) looked at phonological sequences in the acquisition of English by German-speaking children and found three distinct stages: pre-systemic (knowing the sounds only in individual words), transfer (systematically using the L1 categories in the L2), and approximative (restructuring the L2 sounds into a new system). Wolfe Quintero (1992) found that the stages of acquisition of English relative clauses by Japanese learners could be seen as an interaction of six learning strategies.

The answer to the question is that there are indeed sequences of L2 acquisition in different areas of language and with different first languages. The stages through which L2 learners progress have much in common, thus reducing the importance of L1 transfer: Whatever your first language, you'll still go through the same stages in acquiring a second. The question of whether there is an L2 order seems a genuine SLA research question and has received a positive answer.

One problem is how important order is to the study of language acquisition. Chomsky has often spoken of an idealized "instantaneous" model of language acquisition that relates language input and competence as wholes without reference to chronological development (Chomsky 1986). Learning to drive undoubtedly involves stages of mastering the steering wheel, the brakes, the gears, and so on; what matters eventually is whether you can drive, not the stages you

went through: It is the nature of the knowledge of language that is acquired that counts not the accidental stages of acquisition.

A second issue is the definition of stage. Ingram (1989) distinguished, inter alia, continuous stages, plateau stages, and transition stages. Most SLA research such as has used the continuum meaning (Dulay and Burt 1974) apart from processability theory (Pienemann 1998) which is more like a plateau and transition stage (with broad characteristics for each stage.) The fact that second language acquisition proceeds through stages does not in itself mean very much.

Order is then a valid second language acquisition question, given the provisos above. The bulk of SLA research over the last 30 years has certainly shown stagiation in every aspect of second language development.

4 What Are the Similarities between L2 Learning and L1 Acquisition?

A continuing theme has been whether people acquire a second language in the same way as a first. If the L2 stages outlined above are also followed by L1 children, both groups can be presumed to be using the same learning process. The L2 sequence for English grammatical morphemes was similar, though not identical, to that found in L1 acquisition by Brown (1972), the greatest differences being the irregular past tense ("broke"), articles ("the"), copula, and auxiliaries (Dulay, Burt, and Krashen 1982). Other similar sequences of syntactic acquisition have been found in L1 and L2 learning. L2 learners, like L1 learners, start by believing that "John" is the subject of "please" in both "John is easy to please" and "John is eager to please" and only later go on to discover it is the object in "John is easy to please" (Cook 1973; d'Anglejan and Tucker 1975). L2 learners, like L1 children, at first put negative elements at the beginning of the sentence **No the sun shining* and then progress to negation within the sentence **That's no ready* (Wode 1981).

A subtheme underlying several of the questions discussed here is that L1 acquisition is completely successful, L2 learning is not. Take two representative quotations: "Very few L2 learners appear to be fully successful in the way that native speakers are" (Towell and Hawkins 1994: 14); "Relative to native speaker's linguistic competence, learners' interlanguage is deficient by definition" (Kasper and Kellerman 1997: 5). The evidence for this deficiency is held to be the lack of completeness of L2 grammars (Schachter 1988) or the fossilization in L2 learning where the learner cannot progress beyond some particular stage (Selinker 1992), both familiar "facts" in some sense. Part of the interest in SLA research is explaining why L2 learners are mostly unsuccessful. This alleged failure depends upon how success is measured, as we shall see.

The answer to the question of similarities between first and second language acquisition is far from settled. While there are many similarities between L1 and L2 learning, the variation in situation and other factors also produces many differences. One difficulty is filtering out differences that are accidental rather than inevitable. L1 children mostly acquire language in different settings with different exposure to language than L2 learners and they are at different stages of mental and social maturity. It may be inherently impossible to compare equivalent L1 and L2 learners. A more precise version of this question asks whether adults still have access to Universal Grammar in the mind, discussed below. SLA research has not yet assimilated the modern studies of children learning their first language described in books such as Bloom (2002), Clarke (2003), and O'Grady (2005). It is also true that all the comparisons reflect the assumption that ideally second language acquisition is independent of the first language. Given the existence of the other language in the same person's mind, it would be amazing if second language acquisition were identical to first language acquisition.

5 Does Age Affect L2 Learning?

Following on from the last question, while everybody learns their first language with equal ease, there appear to be vast differences between L2 learners, some acquiring a high level of fluency, others a few stumbling words. Such individual differences may reveal crucial aspects of L2 acquisition; the learners' motivation, their cognitive style, their personality, and other individual factors may make a crucial difference. In general Lambert (1990) made a broad distinction between "additive" L2 learning, which adds new skills and experiences to the L2 users' lives, and "subtractive" L2 learning, which detracts from their present state by, say, making them ashamed of their first language. More detailed investigations into differences between L2 learners belong more to psychology than linguistics and are covered in, for example, Dornyei (2006).

The individual factor that has been thought to affect L2 learning most has, however, been the learner's age. There is a universal folk belief, shared by many linguists, that children are better at learning second languages. The academic starting point was Lenneberg's critical period hypothesis (CPH) which claimed language may be learned only within a particular window of opportunity between two months and 13 years of age (Lenneberg 1967), though Lenneberg himself did not extend the CPH directly to L2 learning; successful second language acquisition consists essentially of the achievement of native-like proficiency (Hyltenstam and Abrahamsson 2003). A survey of the CPH can be found in DeKeyser and Larson-Hall (2005). The classic case showing the failure of late L2 acquisition is Joseph Conrad, who wrote his novels in English though born in Poland (Lieberman 1984). Yet, according to Bertrand Russell, he spoke English with "a very strong foreign accent." There are, however, problems with this actual example: Conrad's level of writing in English is clearly exceptional; English was his third language and French, his second, which he learned after the age of 17, was spoken with "elegance" and "no trace of an accent" (Page 1986).

Some research has challenged the superiority of children. If all differences between children and adults are discounted other than age, adults often appear to do better than children. To take a naturalistic learning situation, Snow and Hoefnagel-Höhle (1978) studied how English-speaking people of different ages developed Dutch during their first year in the Netherlands. Adults and 12–15-year-olds outshone younger children aged 3–10 during the first few months; by the end of the year the most successful learners were those between 8 and 10 and the least successful the 3–5-year-olds. Thus, while there seem some advantages for children, adults also have some short-lived assets. This appears also in the classroom situation. Asher and Price (1967) taught Russian through the total physical response method for six weeks to adults and children aged 8, 10, and 14; the adults learned best, the youngest children worst. Yet there is also ample evidence that younger L2 immigrants achieve a higher level in a second language than adults on many tasks, whether accent (Oyama 1976) or communicative abilities (Patkowski 1980).

The accepted wisdom on age has become the view originally summarized in Krashen, Scarcella, and Long (1982) and approved in the massive survey in Singleton and Ryan (2004): Adults acquire second languages better over the short term, children over the long term. In other words, the eventual attainment of children is better even if they start off more slowly. Nevertheless there are problems with the methodology. The learners are often drawn from a limited range of L1s acquiring English in the USA, untypical of the majority of L2 learners in the world. Age is usually defined in terms, not of when the person started learning the L2, but of when they immigrated to a country: Older immigrants inter alia tend to be better educated (Khanna et al. 1998).

The research has little to say directly about acquisition of second languages by children in countries where the second language is not spoken. That is to say, it contributes little to the

perennial educational debate on the best age to start *teaching* a second language. Eckstrand (1976) looked at 2,000 children learning L2 Swedish in schools after their first year and found that the older children were better. The British experience suggested that teaching French in the primary school was not of great help in developing French in the secondary school (Burstall et al. 1974) but this may have more to do with the inability of secondary schools in England to handle children who have already studied French.

The controversy over age still continues. Bongaerts, Planken, and Schils (1995) elicited speech samples from adult native speakers of English and from two adult groups of Dutch L2 learners, one group of 10 people believed to be native-like, the other of 12 who were not. Judges put the top Dutch group within the range of the native speakers, with four Dutch people exceeding them. They argue that "there appear to be cases of 'late' second language learners who can pass for native speakers phonologically," contrary to the belief that a native accent is never attained. Bongaerts, Planken, and Schils (1997) redid the same experiment with tighter controls but still found a proportion of L2 users who were within the bounds of the native group. If an L2 user can pass as a native speaker for a brief period of time, there cannot be any difference between them, just as in the Turing test a computer that cannot be distinguished from a human being has to be credited with intelligence. Piller (2002) has extended this to bilingual couples who often feel that they can pass for native for a particular period of time. This line of research disproves the absolute claim that nobody learns a second language to native level when starting as an adult, at least so far as accent is concerned, however small the proportion of native-like L2 users may be.

But accent is only one aspect of language – does Conrad's foreign accent really outweigh his exceptional command of written English? Research has furthermore relied mostly on the comparison with the accent of the native speaker. In every language accent is used as a way of identifying speakers in terms of status, region, age, sex, and so on. Even L2 learners rapidly learn the appropriate pronunciations for their own gender, for instance that English-speaking men tend to pronounce the 'ing' ending of the continuous form "going" as /in/ but women tend to use /iŋ/ (Adamson and Regan 1991). In a sense no one objects to people from Edinburgh sounding as if they come from Scotland but everyone believes that a Frenchman speaking English should not sound as if they come from Paris; native speakers may give away their origins but foreigners mustn't. Non-native speakers have every right to agree with a French winemaker, "My English is not good but my French accent is perfect." Accent may be a misleading attribute for age effects. Vocabulary research that studied English learners of French below or above the age of 12, for instance, found no differences in the acquisition of vocabulary (Singleton and Ryan 2004).

Again the main interest lies not so much in the data as in the explanations. Age manifests itself as change in the user or the user's environment in some way; the question is which of these changes affects L2 learning. Diverse explanations are offered ranging from changes in brain chemistry (Pulvermüller and Schumann 1994) to a shift in speech processing toward categorisation at about the age of seven (Flege 1992) to the unavailability of Universal Grammar, to be discussed below. Age does seem to have effects on L2 learning but their exact nature is unclear and their causes are mostly speculative.

The SLA research question about the effects of age on language learning has not produced conclusive results after several generations of research. Mostly it has been restricted to immigrants to English-speaking countries and has concerned grammaticality judgments or certain structuralist views of syntax rather than a range of languages and learners. Yet it still continues to attract a flood of research. There are now perhaps few people who deny that age of acquisition has an effect on second language acquisition so far as approximation to native speaker-like language is concerned. Whether this is a valid question for SLA research hinges on the target of native-like speech that is invariably assumed, to be discussed below. It is not just that the measures that are used are taken from native speakers; it is that L2 users are seen as defective when they do not speak like natives The age-related research becomes impossible to interpret

once this target is removed, leading to the question of what successful L2 users know discussed in the next section.

6 Do L2 Learners Attain the Same Level of Language as Native Speakers?

The question of the endpoint of L2 acquisition was already implicit in the questions about the initial state and age but has been raised more explicitly in recent years: What is the final state that L2 users can reach in the knowledge of a second language? Despite the interlanguage assumption that L2 learners have independent grammars, the final state of the L2 learner has frequently been seen in terms of whether L2 learners can achieve the same competence as a native, often called "ultimate attainment."

The starting point was a study by Coppieters (1987), who gave grammaticality judgements to near-native and native speakers of French on nine syntactic structures. Though the near-natives hardly deviated from the native speakers on some structures, on others they differed more, for example, tense contrasts. Even these advanced L2 users could therefore still be distinguished from native speakers. Their ultimate attainment differed from that of the native speaker.

Birdsong (1992) criticized the Coppieters research on a number of counts and essentially redid the experiment with near-native speakers of French with English as L1 and native speakers. He found that, while it was true that the near-natives differed from the natives as a group, when treated as individuals 15 of the 20 near-natives were within the native speaker range, while in the Coppieters study none were. That is in effect five people should not have formed part of the near-native group. In the case of these speakers L2 attainment did not differ from that of native speakers.

White and Genesee (1996) continued this approach by comparing native speakers of English and L2 learners, divided into near-native and non-native groups, who were given a timed grammaticality judgements test of questions such as *Which one are you reading a book about?* and **Who did you meet Tom after you saw?* There were no differences between the natives and near-natives in accuracy and speed, with the exception of sentences such as *Which movies do the children want to rent?* The conclusion is that "Ultimate attainment in an L2 can indeed be native-like in the UG domain" (White and Genesee 1996: 258).

The balance of the research to date suggests that a small proportion of L2 learners can acquire the same knowledge of a language as native speakers, just as a small group seem able to acquire a native-like accent. But the question remains whether closeness to the native speaker is an appropriate yardstick to measure them by. Birdsong (1992: 717) construes "ultimate attainment in L2A [second language acquisition] in terms of whether non-natives can display evidence of possessing native linguistic norms." But bilinguals use languages for different purposes from monolinguals and have a total language system of far greater complexity in their minds: Why should L2 users be measured against the knowledge of a person with only one language? As Sridhar and Sridhar (1986) point out, "Paradoxical as it may seem, Second Language Acquisition researchers seem to have neglected the fact that the goal of SLA is bilingualism." Indeed it is evident that L2 users can become more proficient than average L1 users, as we saw with Conrad's writing. L2 users, for instance, make fewer spelling mistakes in English than 15-year-old native children. Relating the L2 ultimate state to native speakers may be convenient but does an injustice to the overwhelming majority of L2 users, who are thereby seen as failures for not achieving something which is, by definition, not achievable. The unique status of the two languages of the L2 user has been abandoned in favor of seeing whether the L2 user is a defective version of the L1 user. A true SLA research question would be the neutral "What is the final competence of L2 users?" not the monolingual-biased "Can L2 users speak the same as native speakers?"

7 How Important Is Transfer to L2 Learning?

Transfer from the first to the second language involves both use and acquisition, i.e., it may affect both the processes of speaking in the short term and the processes of learning over a period of time. The influence of the first language on the second is obvious from our everyday experience; most native speakers of English can tell whether an L2 user comes from Japan, Germany, France, or Spain.

Some early research, however, attempted to minimize the role of L1 influence. Grammatical morphemes research, for example, suggested that people with different L1s had similar acquisition sequences (Krashen, Scarcella, and Long 1982). Dulay and Burt (1974) tried to quantify transfer mistakes *vis-à-vis* developmental mistakes, claiming that only 24 out of 513 mistakes by Spanish learners of English could be ascribed to their L1.

In the days when linguists considered all languages varied from each other in arbitrary ways, each pair of languages had to be compared from scratch through contrastive analysis. Now that most linguists are concerned with overall relationships between languages, transfer can be seen to utilize overall systematic relationships between languages. Take the example of writing systems. These are mostly held to fall into two groups: meaning-based systems as in Chinese characters and sound-based systems as in the alphabetic system used for English (Paap, Noel, and Johansen 1992). L1 transfer involves carrying the characteristics of the L1 writing system over to the L2. Chinese L1 speakers acquiring the Japanese syllabic writing system (*kana*) rely more on visual strategies, English users on phonological strategies (Chikamatsu 1996); Chinese L1 students have difficulty processing nonwords in English, showing their phonological processing is underdeveloped (Holm and Dodd 1996). Speakers with meaning-based L1 writing systems are better at visual reading tasks in English than those with sound-based L1s (Brown and Haynes 1985). As in other areas, L1 transfer can be a help as well as a hindrance.

Other writing system research has looked at L1 transfer in spelling. Adult Spanish learners of English show characteristic Spanish transfer mistakes involving the double letters <rr> and <ll> and transpositions involving <l> or <r> (Bebout 1985). 38.5 percent of English spelling mistakes made by ten-year-old Welsh/English bilingual children can be attributed to interference from Welsh, whether from phonological interference in the L2 pronunciation, orthographic interference from Welsh sound/letter rules, or transfer of cognate words (James et al. 1993). Different L1s produce characteristic spelling mistakes in English; Japanese learners of English frequently confuse <l> and <r> as in **walmer*, **grobal*, and **sarary*, perhaps because of their well-known pronunciation difficulties with the sounds /l/ and /r/, perhaps because of the way that English loanwords are spelled in the kana syllabic system in Japanese.

Research into phonological transfer has also progressed from lists of phonemes to more general aspects. In the acquisition of English stress assignment by speakers of Polish and Hungarian, 95 percent of the mistakes consisted of transfer of L1 metrical settings (Archibald 1993). English syllables are made to conform to the structure of the L1 by adding epenthetic syllables – **filoor* (Egyptian "floor"), **piliz* (Hindi "please"), and **iskul* (Bengali "school") (Broselow 1992). The role of transfer may change during L2 development. Major (2001, 2002) claims that phonological transfer decreases over time while developmental factors first increase, then decrease.

The transfer of pragmatic speech functions from L1 to L2 has mostly been seen negatively. German learners of English produce requests that are too direct (Kasper 1981); L2 learners of English thank people in ways that are more formal than native speakers, "Thank you very much" rather than "Thanks" (Cook 1985b). Again research has gone from unique features of languages to universal schemes. There is an overall pattern to apologizing in any language consisting of explicit apology, explanations, denial of responsibility, and so on, varying in weight and emphasis from one language to another (Kasper 1996; Bergman and Kasper 1993). An inappropriate

linguistic form may be transferred; a learner may use the conventional Japanese way of refusing by making a statement of principle: "I never yield to temptations" (Beebe, Takahashi, and Uliss-Weltz 1990), with odd effects in English.

Transfer has been looked at within the competition model (MacWhinney 1987). This claims that all languages make use of four cues for the subject of the sentence – word order, agreement, case, and animacy – but these are weighted differently across languages. Thus German speakers should rely on the agreement between verb and subject **The horse hits the cow*; Italian speakers on the subject coming first **The lamb a dog pats*; and Japanese on the subject being animate **The eraser the cow kisses*. L2 users indeed tend to carry over the weightings from the first language and only gradually lose them over time, whether Japanese animacy affecting English (Harrington 1987), or Dutch agreement affecting English (Kilborn and Cooreman 1987): Issidorides and Hulstijn (1992), however, showed that animacy may be an over-riding factor with both English and Turkish learners of Dutch. Transfer here is carrying over the L1 weightings for processing the sentence to the L2.

The interpretation of transfer within the Universal Grammar (UG) theory has taken the most general point of view. If both L1 and L2 represent different choices from the same possibilities laid down by UG, the question of transfer is whether the L1 choices carry over into the L2 knowledge. The principles and parameters model of UG has the great advantage of providing an overall descriptive syntactic model within which the two languages can be compared, even if UG theory changes constantly, called by Rizzi (2004) "comparative syntax." It will be discussed in the next question.

In general transfer has become a less overt concern in SLA research and has been subsumed within other issues concerned with the relationship between the two languages. Weinreich's original definition indeed allowed transfer to go in both directions: "Those instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language" (Weinreich 1953: 1). The second language may also have an effect on the user's first language. In phonology this has been a popular subject of investigation. Even though two languages seem to have the same phoneme, this may disguise differences in voice onset time (VOT). L2 research has shown that French learners of English have a longer VOT for the voiceless /t/ sound in their L1 French than monolingual speakers (Flege 1987). Effects are also found in intonation: Dutch L2 users of Greek have slightly different question intonation from monolinguals (Mennen 2004); German children who speak Turkish have different intonation from those who don't (Queen 2001). Similarly L1 meanings for words may be influenced by the L2; a monolingual speaker of Korean uses "paran sekj" (blue) to mean something greener and less purple than a Korean who also knows English (Caskey-Sirmons and Hickerson 1977). Even language functions transfer from L2 to L1 (Kasper 1996); Locastro (1987), for example, found English speakers of Japanese using "aizuchi" (nodding for agreement) when talking English.

Transfer in the sense of the relationship between the two languages in the same mind is crucial to second language acquisition research. If people simply acquired a second language in the same way as their first, there would be no need for a separate discipline of SLA research. A major factor in the different courses of L1 and L2 acquisition must be the developing links between the two languages. In a sense any investigation of L2 learning or use that does *not* involve this relationship is not SLA research. The word transfer "tends" now to be avoided and less model-specific terms like "cross-linguistic influence" (CLI) preferred (Kellerman and Sharwood-Smith 1986). Hence modern treatments not only talk of L1 to L2 "forward" transfer but also of L2 to L1 "reverse" transfer, L3 to L2 "lateral" transfer, and "bidirectional" transfer in which any two languages known by the speaker affect each other reciprocally (Jarvis and Pavlenko 2008). The research question of how different languages relate in the L2 user's mind is the unique heart of SLA research.

8 What Is the Relationship between Universal Grammar and Second Language Acquisition?

The Universal Grammar (UG) model of language acquisition developed in the 1980s, called principles and parameters theory, claims that the child's mind possesses universal principles that always apply to language and variable parameters that have different settings in different languages (Chomsky 1986); this insight survives intact in the modern proposals for the Minimalist Program (Chomsky 1995). A sentence such as **Is Sam is the cat that black?* is not only impossible in English but is also forbidden in *any* human language because elements in the sentence can only be moved around to form questions according to the structure of the sentence, not its linear order (Chomsky 1980); this principle of structure-dependency is built in to the human mind so that a human language that breaks structure-dependency is literally inconceivable.

Sentences must have subjects in English "He speaks" not **Speaks*, French "Il parle," and German "Er spricht," but they are not compulsory in Italian "Parla," Japanese "Yomu," or Arabic "Yatakallamu." The claim is that a language belongs either to the group of pro-drop languages that permit null subjects or to the group of nonpro-drop languages that have compulsory subjects – the pro-drop or null subject parameter. L1 children do not need to learn the principles of UG because they are invariably true – no language could possibly break structure-dependency. But children do need to set the parameters appropriately for the language they are acquiring, say to pro-drop or nonpro-drop, which means hearing the right language input to set them. And they also need to acquire a vast store of lexical information about how words behave within the structure of the sentence.

The main question has been whether L2 learners have access to Universal Grammar or have to manage without it; in other words, is L2 learning like L1 learning? This is often phrased as a choice between direct access to UG, indirect access, and no access (Cook 1985a). In direct access, the L2 learner applies the mental faculty of UG to the L2 input without hindrance and acquires a grammar consisting of the same principles, parameter settings, and so on, as the L1 speaker: L2 learning is just like L1 acquisition. In indirect access, the L2 learner is able to access UG only via the L1 knowledge in the mind; those parts of UG that have been activated in the first language can be used again, but other parts are not available. In no access, the L2 learner is effectively cut off from UG; everything has to be learned using other aspects of the mind.

The argument for direct access often recapitulates one used in L1 acquisition: If L2 learners know something they could not have learned from L2 input or from their L1 knowledge, it could only come from the UG in their minds. In L1 acquisition, structure-dependency is part of UG because it could not be acquired from input (Chomsky 1988). However, languages which do not form questions etc. by movement do not need structure-dependency. If speakers of such L1s learn structure-dependency in a second language, this knowledge could not derive either from input or from their first language but must come directly from UG itself. Japanese (nonmovement) learners of English indeed know structure-dependency according to Otsu and Naoi (1986); a range of L2 learners of English, including Japanese, Chinese, and Finnish L1s, all scored more than 86 percent on a test of structure-dependency (Cook 2003b). So L2 learners clearly have direct access to UG, at least so far as this principle is concerned.

The evidence for indirect access to UG is the effects of L1 parameter settings on the L2, i.e., a version of transfer. Japanese and Spanish learners, for example, are influenced by the word order preferences of their first language in interpreting English sentences (Flynn 1987). The pro-drop parameter concerning the compulsory presence of subjects in the sentence has been massively studied. French learners of English, with the same nonpro-drop setting in L1 and L2, were much better at saying that **In winter snows a lot in Canada* was ungrammatical than Spanish learners, who have a pro-drop setting in the L1 (White 1986); both English and French learners of Spanish had, however, no problems with acquiring the Spanish pro-drop setting despite their different

L1 settings (Liceras 1989). Later research has linked the presence of subjects to verb inflections in the present tense, with confusing results (Lakshmanan 1991; Hilles 1991). Much research with parameters has tended to show effects from the L1, that is to say L2 learners' access to UG is filtered indirectly through their first language.

The test for access shifted to the presence or absence of verb and noun inflections such as past tense 'ed' ("he liked") and agreement '-s' ("he likes") in L2 learners' speech. If L2 learners show a lack of functional phrases at the outset, as found in L1 children (Radford 1990), this could confirm the direct access position. Early SLA research had already in a sense made a similar point by demonstrating the absence of grammatical morphemes in L2 learners. While Meisel and Müller (1992) found these functional phrases in early L2 acquisition of German and Grondin and White (1996) found them in the early stages of L2 acquisition of French, Vainikka and Young-Scholten (1991) suggested that the sentences of early L2 learners of German include only the parts of the Verb Phrase, that is to say they are not really functional phrases after all.

Finally, the no access position has largely based itself on two propositions. One is that L2 learning could use other mental faculties than UG, such as general problem-solving abilities; this might be plausible if concrete suggestions were spelled out rather than simply mentioned. The other is that L2 learners do not acquire the L2 as well as the native speaker (Schachter 1988; Bley-Vroman 1989), summarized earlier. While L2 learners indeed score less than natives on most UG-related syntax tests, they also score less on other tests of cognitive functioning introduced via the L2, such as working memory (Brown and Hulme 1992) – the so-called L2 cognitive deficit. Such "deficiencies" may not be part of UG itself but reflect the overall working of their cognitive apparatus.

Deciding whether learners have access to UG is fraught with difficulties. Different research methods and different syntactic analyses come up with conflicting answers with no clear way of reconciling them; indeed the UG theory itself changes so rapidly that principles have a half-life of about five years. Access may not be a real question because it reifies UG as a separate object that learning has access to rather than as the changing state of the mind itself. The relationship of L2 learning to L1 learning has been left problematic. There seem strong similarities, but there are also differences, perhaps due to the greater maturity of most L2 learners causing social or cognitive differences not directly part of language learning. The crucial SLA research question is whether the L2 learner's final state of language knowledge fits UG, not whether it fits native speaker grammars.

The underlying question for UG researchers has however shifted toward the nature of the initial state of the L2 learner's mind, expressed as a series of hypotheses (Cook and Newson 2007):

- *the L2 learner has no Universal Grammar to build on*. Evidence for this is the same as for no access above.
- *the L2 learner has a second copy of Universal Grammar to build on*, called the full access hypothesis. According to its proponents UG is "continuously available to assist in the construction of various language-specific grammars" (Flynn and Lust 2002: 98).
- *the L2 learner can build on Universal Grammar in so much as it has been incorporated into the first language knowledge*, known as the full access/full transfer hypothesis. This was advanced on the basis of one Turkish learner of L2 German who showed both influence from L1 Turkish and restructuring based on UG (Schwartz and Sprouse 1996).
- *the L2 learner can partly build on UG*, i.e., it is only partially available, a partial access hypothesis. variants of this include the minimal trees hypothesis that functional projections develop out of the lexical categories used in early acquisition (Vainnikka and Young-Scholten 1996) and the failed functional features hypothesis that only the UG features instantiated in first language acquisition are available to the L2 learner (Hawkins and Chen 1997).

Much SLA research in this so-called "generative" approach has then developed into "hypotheses wars" over the initial state. While it may perhaps be possible to extrapolate from such

small evidence to all L2 learners everywhere, it reflects a limited view of the richness of ideas about syntax and language acquisition of the Universal Grammar tradition. While it is showing a unique fusion of current syntax from the Chomskyan tradition with second language acquisition, it has become detached from the main body of SLA research and its tendency to jump on the latest syntactic bandwagon means that most of its results are superseded in a year or two rather than building into a solid edifice.

9 What Is the Role of Language Input?

Everyone would agree that people do not learn an L2 if they encounter no examples of it. Beyond this, there is little agreement over the role that language input plays, the amount that is necessary, and the form that it should take. Perhaps the differences between L1 and L2 learning or between learners of different ages stem only from the different types of language they encounter. If certain types of input are more effective than others, this would have dramatic consequences for language teaching.

The starting point for research was the characteristics of language addressed to children. In many languages, adults use baby talk to children, not only peculiar words such as "moomoo" in Japanese, "baâ" in Arabic, and "moocow" in English but also a higher frequency of commands and questions (Newport 1976). Is there an equivalent foreigner talk to L2 users? Freed (1981) found that the types of sentence addressed to L2 users were more like that used to native adults than to children, reflecting the different topics adults talk about. Foreigners asking the way on the streets of Wellington were not, however, addressed differently from natives (Stocker-Edel 1977).

The issue then shifted, as it had done in L1 acquisition research, away from the grammatical features of the language input to the interaction between learner and nonlearner. For example, while it is true that non-native speakers are addressed differently in places such as travel agents, this is mostly due to the information they receive being more low level, such as "It's a big jet" in response to "What kind of plane is it?" (Arthur et al. 1980). Giving students opportunities to interact improved performance compared with editing or simplifying the language they heard (Pica, Young, and Doughty 1987).

In the UG theory, some language input is necessary in acquisition in order to set the parameters and to acquire vocabulary. Everything L1 children need must either come from their minds or be present in the input as "positive evidence" – sentences that they hear – rather than "negative evidence" – parents supplying corrections or pointing out forms that do not occur. Parameter-setting in L2 as in L1 requires responding to features of the input that the learner can make out. To set the pro-drop parameter, for example, it may be necessary to hear all the forms of the present tense.

A more radical view is that negative evidence is needed in L2 acquisition even if irrelevant to L1 acquisition. Possibly the L1 has put the learner into a position that is irretrievable from positive evidence alone – a highly restrictive version of indirect access. Negative evidence in an L2 context might be the teachers' corrections of students' speech or explanations that give the learners information about the facts of the language. French allows an adverb between a verb and a direct object, "Jean embrasse souvent Marie," but English does not, **John kisses often Mary*. White (1991) successfully taught English learners of French where to place adverbs in the sentence, thus using negative evidence to overcome their L1 parameter setting. Conversely Trahey and White (1993) exposed French-speaking children acquiring English to an input "flood" of English containing adverbs, leading to an increase in preverbal adverbs, "Anna carefully drives her new car," but not to the decline of the ungrammatical postverbal adverbs **Anna drives carefully her new car*. Thus negative evidence in the form of explanation can play an important role in L2 acquisition. Indeed, it may be possible to enhance the L2 input to highlight specific points (Sharwood-Smith 1993), as is indeed claimed for phonological clues in speech addressed to L1 children (Morgan 1986).

Another form that input takes is interaction: How people interact with learners may help them acquire the language. In the interaction approach "negotiation of meaning" is crucial: "the process in which, in an effort to communicate, learners and competent speakers provide and interpret signals of their own and their interlocutor's perceived comprehension" (Long 1996: 418). L2 learners develop by trying to get things across to others and by getting feedback on their performance (Lyster and Ranta 1997). This may too reflect a monolingual bias to research: Macaro (2005: 66) argues that the "unswerving faith in the comprehensible input – negotiation – comprehensible output has been entirely due to the fact that the proponents of these theories and hypotheses simply did not speak the first language of their subjects or students." An alternative view of interaction is provided by the idea of zone of proximal development (ZPD) based on the theories of the early twentieth-century Soviet psychologist Vygotsky (Vygotsky 1962 [1934]). The ZPD is defined by him as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in cooperation with more capable peers" (Vygotsky 1978 [1935]: 86). SLA research usually interprets this through the idea of scaffolding – the process of social assistance through which the learner is assisted to get to the next point in development by other people (Lantolf 2000).

So, in general, language input and interaction seem to play a similar role in L2 learning, apart perhaps from the need for some negative evidence. This may be useful to language teaching whose main influence on students is in a sense controlling their experience of the L2.

10 What Strategies and Processes Do L2 Learners Use?

The term "strategy" has been applied in L2 research to the mental processes, conscious or otherwise, used by L2 learners for learning and communication, often relying on theories from psychology such as the adaptive control of thought (ACT*) model from Anderson (1983) or Levelt (1989). Much of it concentrates on compiling lists of strategies from observation rather than examining data from recordings or from experiments.

Early research into good language learner strategies tried to isolate the processes used by successful L2 learners. Extensive research in Canada found that good learners tend to adopt the learning style that suits them, to involve themselves in the language learning process, and so on (Naiman et al. 1995), called by McDonough (1995) "wholesome attitudes" rather than strategies.

The research summarized in O'Malley and Chamot (1989) focused more on learning, dividing strategies into "metacognitive strategies" for managing thinking, such as monitoring one's speech, "cognitive strategies" for thinking itself, such as note-taking, and "social strategies" which involve other people, such as asking for help. When O'Malley and Chamot (1989) asked students to report what they used, a cumulative list of 27 such strategies emerged. Most were nonlinguistic in that they could apply equally to the learning of any subject rather than being unique to L2 learning. Such strategies say more about the characteristics of academic students in formal classrooms than about L2 acquisition itself.

A distinct branch of strategy research has concerned vocabulary. Cohen (1990), for instance, showed that students remember vocabulary best when they make associations and learn cognates. Often this became linked to the idea of mnemonic strategies used since the ancient Greeks in which vocabulary items are associated with already memorized key words (Paivio and Desrochers 1979) or vivid images tie the new word into existing memories, summarized in Nation (2001). It can be disputed, however, whether vocabulary learned in this fashion is readily used for everyday language purposes, as opposed to being produced in language tests or exercises of an artificial type.

The area of communication strategies had an easier task since the success of a communication strategy can be more readily gauged. In the L2 literature a communication strategy is needed

only when things go wrong – a spare tyre for when your car has a puncture rather than the steering wheel you use all the time. Lists of communication strategy were devised using categories such as "approximation" "animal" for "horse" (Tarone 1980) or "literal translation" "green things" for "vegetables" in a literal translation from Danish (Faerch and Kasper 1983).

However, the investigation of communication strategies in actual use by Poulisse (1990, 1996) showed not only that the majority of such strategies reflect a lack of vocabulary rather than of grammar, but also that they are used in the first language when speakers lack the vocabulary to express what they want to say. Hence they are better called "compensatory strategies" since they fill in gaps in vocabulary whether in the first or the second language. Thus the idea of communication strategy became part of normal language use rather than a specifically L2 phenomenon.

The underlying issue is whether strategies are born afresh in an L2 or are carried over from the speaker's existing knowledge. Strategies are a reminder that L2 learners bring more to L2 learning than L1 grammatical competence and that they need to communicate effectively in both languages by whatever means they can. Some of the methodological problems with strategies research have been highlighted by Macaro (2006).

11 How Are the Two Languages Related in the Mind?

To some extent we can ask whether L2 users comprehend and produce speech in similar or different ways compared to L1 users. The L2 user, however, possesses the unique process called code-switching in which the speaker changes language in midstream, sometimes between sentences but often within the bounds of the same sentence, as in *Suami saya dulu slim and trim tapi sekarang plump like drum* (Before my husband was slim and trim but now he is plump like a drum) produced by a Bahasa Malaysia/English speaker. Grosjean (1989) sees L2 users as having two modes of language use: one is the monolingual mode in which one or other of the two languages is employed, the other is the bilingual mode in which both languages are used at once. Mostly code-switching research has related to language use in advanced bilinguals, not to how learners code-switch in the early stages or within the classroom.

Code-switching within a single sentence can be investigated in terms of the points in the syntactic structure where a switch can take place. Poplack (1980) proposed two constraints. The "free morpheme constraint" is that the speaker may not switch language between a word and its inflection unless the word is pronounced as if it were in the language of the ending; hence it is possible to have an English/Spanish switch *flipeando* (English "flip" + Spanish "ando") as "flip" is possible in Spanish but not **runeando* as "run" is impossible. The "equivalence constraint" is that the switch-point must not violate the grammar of either language; so it is possible to have the English/French switch *J'ai acheté an American car* as it preserves the grammar of both languages but not to have **a car americaine* as this would violate English word order.

Other models of code-switching have relied on deeper syntactic analysis. The "government" model of code-switching proposed that the switch cannot come within a maximal phrase (DiSciullo, Muysken, and Singh 1986), that is to say a lexical head of a phrase forces the rest of the phrase into the same language; for example, the head "see" governs the object Noun Phrase in "see the book" and so keeps the rest of the phrase "the book" in English.

The alternative 4M Model (Myers-Scotton 2005) sees code-switching in terms of a matrix language and an embedded language. Morphemes are divided into four types: content morphemes, early system morphemes, late bridge system morphemes, and late outsider system morphemes. Content and early system morphemes go with the embedded language in depending on meaning. The late bridge and outsider system morphemes go with the matrix language as they provide the grammatical framework within which the content and early system morphemes can be placed.

Exceptions have been found to all of these constraints. For example the word "aunties" is found in Panjabi/English switches both with the English ending /a:nti:z/ and with the Panjabi ending /a:ntija ã/ (Gardner-Chloros 1995), despite the free morpheme constraint. Arabic to Dutch code-switches take place between indirect and direct object *žib li-ya een glas water of zo* (Get for-me a glass of water or so) (Muysken 1995), despite being within the same verb phrase. The constraints seem probabilistic rather than determinative.

Code-switching can be distinguished from other uses of L1 vocabulary in word borrowing or communication strategies by its functions. Overall, code-switching is used by one speaker when the other participant knows both languages rather than resorted to out of ignorance of some aspect of the L2. Romaine (1994) list reasons such as: reporting speech, appropriacy of topic, directing speech at one of several people present, and emphasizing the status of one language or the other.

Such lists are open ended and do not reflect the switching conventions between two given languages in a particular situation. Again the question is whether codeswitching involves two languages at once or fits some elements of one language within the framework of another. Investigating how a speaker can use two language systems at once reveals how the two languages relate in the same mind rather than how they work separately. In a sense the fundamental issue is still whether the knowledge of the two languages in one mind is separate or combined, back to the starting point of Weinreich (1953). The two languages coexist in the same mind: Bilinguals do not have two heads. Yet clearly L2 users can separate the two languages, consciously choosing which language to use. At some level the two languages must be distinct.

This choice has often featured in the debate about whether bilingual children have one language system or two. A common talking point is whether children mix the vocabulary of their two languages. Two children learning Italian and German had a single lexical system which separated at the next stage into two lexicons with one syntax (Taeschner 1983). Early sentences with unequal proportions from the vocabulary of the two languages have also been found such as *Quelle couleur your poupée?* (Swain and Wesche 1975). A child learning English and Dutch used only 4.3 percent mixed sentences in speech addressed to Dutch speakers, 3.9 percent to English speakers, and 2.5 and 0.9 percent respectively that were balanced between the two languages ("Dutlish") (De Houwer 1990). Genesee, Nicoladis, and Paradis (1995) found five children under 26 months could differentiate the two languages, even if they code-mixed.

In terms of pronunciation, Leopold (1947) reported some confusion in the speech of his daughter between the sounds of the two languages and some carry-over of phonological processes from one language to the other; Fantini (1985) and Burling (1959) described children in whom the phonology of one language is dominant. Oksaar (1970), however, studied a child who kept the pronunciation of the two languages separate. Schnitzer and Krasinski (1997) found a bilingual Spanish/English child formed a single phonological system before separating the phonologies of the two languages. At best these pieces of research provide counterinstances to any absolute claim that *all* bilingual children necessarily have either merged or separate pronunciation systems.

In syntax, research into early bilingualism such as Burling (1959) found few signs of syntactic interference between the two languages. Dutch and English gender are kept distinct in the child studied by de Houwer (1990); bound morphemes such as the plural stay in one language; the child mostly uses the appropriate Object Verb word order in Dutch and Verb Object in English. Swain and Wesche (1975) find some interference between the two languages, such as the occurrence of French structures in English sentences, for example *They open, the windows?* Others have described a stage when children have two lexicons but a single syntactic system (Volterra and Taeschner 1978). Meisel (1990: 21) concludes that "fusion is not necessarily a characteristic of bilingual language development, but mixing may occur until code-switching is firmly established as a strategy of bilingual pragmatic competence." Paradis and Genesee (1996) found that bilingual children developed English and French functional categories at the same rate as

monolingual children. The consensus seems to be that, after an initial semantically organized phase, children keep the systems of the two languages distinct. Genesee (2002) showed that from an early age they are adept at code-switching to suit the listener. This does not of course mean that the two languages are not part of the same system at some level: Any native speaker uses a variety of styles and register for different purposes; the system I use for handwriting is quite different from the one I use for wordprocessing; there is no way in which I mix them; yet no one would claim that I speak two languages simply because I have two different ways of realizing written English. Indeed, it is very hard to decide what would constitute code-switching in young children, accepted as a normal feature of bilingualism, and what would be code-mixing, apparently undesirable.

This recognition of two distinct systems in young children contrasts to a large extent with research into the phonology of adult L2 users. To go back to the VOT (Voice Onset Time) of stop consonants, bilinguals had about the same VOT whether they were speaking English or Spanish (Williams 1977); English learners of Portuguese lose their L1 VOTs the better they become in the L2 (Major 2001). L2 learners can have a VOT in between the values for the two languages, whichever language they are speaking. L2 users could be thought to have a single system for L1 and L2 phonology, or at least the two systems have influenced each other in both directions at some point in the past. Watson (1991: 44) comes to the conclusion that "the bilingual may have two systems but which differ in some way from those of monolinguals." In reading also, Greeks who know English read Greek differently in some respects, for example, being affected by the order of presentation, while monolinguals are not (Chitiri and Willows 1997).

The evidence for one lexicon or two in adult L2 users is fairly mixed. One line of research has tended to show two separate lexicons: Kirsner (1986) claimed language-specific words had separate representations but not cognates. Another approach emphasises the factors common to the two lexicons, Grosjean (1989), for instance, arguing that the L1 cannot be switched off while processing the L2. A third possibility is that an independent semantic store connects two separate lexicons; Kroll (1993) found that similar meanings carry across languages, not similar forms. Overall the question of whether L2 users have one system or two is no more settled than the other questions despite fairly widespread discussion in early childhood bilingualism, phonology, and vocabulary. Only in syntax has the question hardly arisen, perhaps because, even in the UG model, it is taken for granted that the two grammars are distinct, even if they influence each other in processing or in development.

It is hard to see that much progress has been made in resolving this underlying question of SLA research since the days of Weinreich. There is still conflicting evidence about whether the L2 user has two systems or one and how these relate during development.

Conclusion

Doubtless many other questions and interpretations could be derived from this large and often contradictory area. The field potentially takes in all areas of linguistics and language acquisition, leading to a range of research techniques, described in Tarone, Gass, and Cohen (1994). Many techniques are integral to some subfield of linguistics, such as the phonologist's VOT used in Flege (1987) or the psychologist's task of counting the sounds in words used in Holm and Dodd (1996). Others are borrowed from the mainstream psychology tradition, such as response time measures (White and Genesee 1996), from descriptive linguistics such as analysis of corpora (Klein and Perdue 1992), or from techniques employed with L1 children such as elicited imitation in which learners repeat sentences (Cook 1973; Epstein, Flynn, and Martahardjono 1996). It is hardly surprising that there is a lack of comparability between results and a lack of agreement over conclusions, even when tackling a similar range of issues such as the ten questions seen here. A question such as age may be posed and answered in one

way by specialists in phonology such as Flege (1992), in another by specialists in UG such as Bley-Vroman (1989), both of them legitimate in terms of their own fields of reference. One dangerous technique is the grammaticality judgments task (Birdsong 1989). People are asked to judge whether sentences such as *What did the teacher know the fact that Janet liked?* are grammatical. One problem is the difficulty in connecting such judgments with the speaker's normal language use. Another is that L2 users may not treat such tasks in the same way as monolinguals, for example, translating the sentence back into their L1 (Goss, Ying-Hua, and Lantolf 1994), particularly because of their heightened awareness of language itself (Galambos and Goldin-Meadow 1990; Bialystok 1993).

Virtually all the techniques in the research mentioned here involve an overt or covert comparison of L2 learners with native speakers. The native speaker indeed provides a quick measure of comparison. Taken too seriously, however, this yardstick denies the interlanguage assumption by subordinating L2 users to native speakers. Useful as it may be to compare apples with pears, apples inevitably seem to make poor pears, just as it is persistently claimed that L2 learners make poor native speakers.

This account has looked at a selection of the areas where linguistics and SLA research cross paths and has not done justice to many others, for example, the sociolinguistic approach described in Regan (1998). It has avowedly taken the prime goal of SLA research to be finding out how people learn and use second languages. Some researchers, however, see SLA research, not as a subject in its own right, but more as a testbed for theories of linguistics and language acquisition (Cook 1981; Davies 1996; Epstein, Flynn, and Martahardjono 1996). While other disciplines may find it useful to have access to this rich source of data, as yet SLA research has hardly raised a ripple in the construction of linguistic theories, unlike L1 acquisition. One reason is the sheer recalcitrance of this complex field where no data collection is simple and no learner is tabula rasa; claims are too easily rebutted by other methods, other situations, and other combinations of L1 and L2. A second reason is the rapid obsolescence of linguistic theory, particularly UG where the advent of the Minimalist Theory (Chomsky 1995) has undermined most of the prior UG-related SLA research. SLA researchers who attempt to contribute to UG theory seldom make stable discoveries about SLA because the theoretical ground has shifted under their feet before their research is out.

The uniqueness of SLA is indeed the relationship of the two languages in one person. Chomsky (1986) proposed that the first aim of linguistics is to answer the question "What constitutes knowledge of language?" Virtually all linguists and L2 researchers have answered the question in terms of the knowledge of a monolingual. Is knowledge of a second language indeed a pale reflection of a first, or is it something in its own right – multicompetence (Cook 2003a)? Borer (1996: 719) reminded the field that "The first question to be resolved is what is the steady state of L2 acquisition and whether given a constant L1 and a constant L2, this steady state is homogenous across speakers." The questions we have looked at here all circle around this issue and start to provide some inklings of what knowing and using two languages is actually like, something which, after all, is the everyday state of the majority of human beings. Indeed, arguably linguistics should start from the multilingual language user, the normal state of any human being, rather than the monolingual native speaker who is nor representative of the potential of language in the human mind due to an early environment in which they were deprived of a second language (Cook 2009).

REFERENCES

Adamson, H. and Regan, V. (1991). The acquisition of community norms by Asian immigrants learning English as a second language: a preliminary study. *Studies in Second Language Acquisition*, 13(1), 1–22.

Anderson, J. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Archibald, J. (1993). Language learnability and L2 phonology: the acquisition of metrical parameters. Dordrecht: Kluwer.

Arthur, B., Weiner, M., Culver, J., Young, L., and Thomas, D. 1980. The register of impersonal discourse to foreigners: verbal adjustments to foreign accent. In D. Larsen-Freeman (ed.), *Discourse Analysis in Second Language Research* (pp. 111–24). New York: Newbury House.

Asher, J. and Price, B. (1967). The learning strategy of total physical response: some age differences. *Child Development*, 38, 1219–27.

Bebout, L. (1985). An error analysis of misspellings made by learners of English as a first and as a second language. *Journal of Psycholinguistics Research*, 14(6), 569–93.

Beebe, L. M., Takahashi, T., and Uliss-Weltz, R. (1990). Pragmatic transfer in ESL refusals. In R. Scarcella, E. Andersen, and S. Krashen (eds), *Developing Communicative Competence in a Second Language* (pp. 55–73). New York: Newbury House.

Bergman, M. L. and Kasper, G. (1993). Perception and performance in native and non-native apologising. In G. Kasper and S. Blum-Kulka (eds), *Interlanguage Pragmatics* (pp. 82–107). Oxford: Oxford University Press.

Bialystok, E. (1993). Metalinguistic dimensions of bilingual language proficiency. In E. Bialystok (ed.), *Language Processing in Bilingual Children* (pp. 113–40). Cambridge: Cambridge University Press.

Birdsong, D. (1989). *Metalinguistic Performance and Interlanguage Competence*. Berlin: Springer.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, 68, 706–55.

Bley-Vroman, R. W. (1989). The logical problem of second language learning. In S. Gass and J. Schachter (eds), *Linguistic Perspectives on Second Language Acquisition* (pp. 41–68). Cambridge: Cambridge University Press.

Bloom, P. (2002). *How children Learn the meaning of words*. Cambridge, MA: MIT Press

Bloomfield, L. (1933). *Language*. New York: Holt.

Bongaerts, T., Planken, B., and Schils, E. (1995). Can late starters attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton and Z. Lengyel (eds), *The Age Factor in Second Language Acquisition* (pp. 30–50). Clevedon: Multilingual Matters.

Bongaerts, T., Planken, B., and Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, 19, 447–65.

Borer, H. (1996). Access to UG: the real issues. *Behavioral and Brain Sciences*, 19, 718–20.

Broselow, E. (1992). Non-obvious transfer: on predicting epenthesis. In S. Gass and L. Selinker (eds), *Language Transfer in Language Learning* (pp. 71–86). Amsterdam: Benjamins.

Brown, G. D. A. and Hulme, C. (1992). Cognitive processing and second language processing: the role of short term memory. In R. J. Harris (ed.), *Cognitive Processing in Bilinguals* (pp. 105–21). Amsterdam: Elsevier.

Brown, R. (1972). *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.

Brown, T. and Haynes, M. (1985). Literacy background and reading development in a second language. In T. H. Carr (ed.), *The Development of Reading Skills*. Cambridge, MA: Academic Press.

Burling, R. (1959). Language development of a Garo and English speaking child. *Word*, 15, 45–68.

Burstall, C., Jamieson, M., Cohen, S., and Hargreaves, M. (1974). *Primary French in the Balance*. Slough: NFER.

Caskey-Sirmons, L. A. and Hickerson, N. P. (1977). Semantic shift and bilingualism: variation in the colour terms of five languages. *Anthropological Linguistics*, 19/8, 358–67.

Chikamatsu, N. (1996). The effects of L1 orthography on L2 word recognition. *Studies in Second Language Acquisition*, 18, 403–32.

Chitiri, H.-F. and Willows, D. M. (1997). Bilingual word recognition in English and Greek. *Applied Psycholinguistics*, 18, 139–56.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1980). On cognitive structures and their development. In M. Piattelli-Palmarini (ed.), *Language and Learning: The Debate between Jean Piaget and Noam Chomsky* (pp. 35–52). London: Routledge.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.

Chomsky, N. (1988). *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. and Halle, M. (1968.) *The Sound Pattern of English*. New York: Harper and Row.

Clarke, E. V. (2003). *First Language Acquisition*. Cambridge: Cambridge University Press

Cohen, A. (1990). *Language Learning: Insights for Learners, Teachers, and Researchers*. New York: Newbury House/Harper Row.

Cook, V. J. (1973). The comparison of language development in native children and foreign adults. *International Review of Applied Linguistics*, 11(1), 13–28.

Cook, V. J. (1975). Strategies in the comprehension of relative clauses. *Language and Speech*, 18, 204–12.

Cook, V. J. (1981). Some uses for second language learning research. *Annals of the New York Academy of Sciences*, 379, 251–8.

Cook, V. J. (1985a). Chomsky's Universal Grammar and second language learning. *Applied Linguistics*, 6, 1–8.

Cook, V. J. (1985b). Language functions, social factors, and second language teaching. *International Review of Applied Linguistics*, 13(3), 177–96.

Cook, V. J. (ed.) (2003a). *Effects of the Second Language on the First.* Clevedon: Multilingual Matters.

Cook, V. J. (2003b). The poverty-of-the-stimulus argument and structure-dependency in L2 users of English. *International Review of Applied Linguistics*, 41, 201–21.

Cook, V. J. (2007) Classic review: syntactic structures 50 years on. *International Journal of Applied Linguistics*, 17, 1, 120–31.

Cook, V. J. (2008). Review of Kroll, J. and De Groot, A. (eds) *Handbook of Bilingualism*, *Language*, 84(1), 196–9.

Cook, V. J. (2009). Multilingual universal grammar as the norm. In I. Leung (ed.), *Third Language Acquisition* (pp. 55–70). Bristol: Multilingual Matters.

Cook, V. J. and Newson, M. (2007). *Chomsky's Universal Grammar: An Introduction*, 3rd edn. Oxford: Blackwell.

Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language*, 63, 3, 545–73.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 5, 4, 161–70.

Corder, S. P. (1971). Idiosyncratic errors and error analysis. *International Review of Applied Linguistics*, 9, 2, 147–59.

d'Anglejan, A. and Tucker, G. R. (1975). The acquisition of complex English structures by adult learners. *Language Learning*, 15, 2, 281–93.

Davies, W. D. (1996). Morphological uniformity and the null subject parameter in adult SLA. *Studies in Second Language Acquisition*, 18, 475–93.

DeKeyser, R. and Larson-Hall, J. (2005). What does the critical period really mean? In Kroll and De Groot 2005, pp. 88–108.

De Houwer, A. (1990). *The Acquisition of Two Languages from Birth: A Case Study.* Cambridge: Cambridge University Press.

DiSciullo, A. M., Muysken, P., and Singh, R. (1986). Government and code-mixing. *Journal of Linguistics*, 22, 1–24.

Dornyei, Z. (2006). *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition.* Mahwah, NJ: Erlbaum.

Dulay, H. C. and Burt, M. K. (1973). Should we teach children syntax? *language learning*, 3, 245–57.

Dulay, H. C. and Burt, M. K. (1974). Errors and strategies in child second language acquisition. *TESOL Quarterly*, 8, 2, 129–36.

Dulay, H. C., Burt, M. K., and Krashen, S. (1982). *Language Two.* Rowley, MA: Newbury House.

Eckstrand, L. (1976). Age and length of residence as variables related to the adjustment of migrant children with special reference to second language learning. In G. Nickel (ed.), *Proceedings of the Fourth International Congress of Applied Linguistics* (pp. 179–97). Stuttgart: Hochschulverlag.

Ellis, R. (2008). *The Study of Second Language Acquisition.* Oxford: Oxford University Press.

Epstein, S. D., Flynn, S., and Martahardjono, G. (1996). Second language acquisition: theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, 19, 677–758.

Faerch, C. and Kasper, G. (1983). Plans and strategies in foreign language communication. In C. Faerch and G. Kasper (eds), *Strategies in Interlanguage Communication* (pp. 20–60). London: Longman.

Fantini, A. (1985). *Language Acquisition of a Bilingual Child: A Sociolinguistic Perspective.* San Diego, CA: College Hill Press.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47–65.

Flege, J. E. (1992). Speech learning in a second language. In C. Ferguson, L. Menn, and C. Stoel-Garnmon (eds), *Phonological Development: Models, Research, and Applications* (pp. 565–604). Parkton, MD; York Press.

Flynn, S. (1987). *A Parameter-Setting Model of L2 Acquisition.* Dordrecht: Reidel.

Flynn, S. and Lust, B. (2002). A minimalist approach to L2 solves a dilemma of UG. In V. J. Cook (ed.), *Portraits of the L2 User* (pp. 93–120). Clevedon: Multilingual Matters.

Freed, B. (1981). Foreigner talk, baby talk, native talk. *International Journal of the Sociology of Language*, 28, 19–39.

Galambos, S. J. and Goldin-Meadow, S. (1990). The effects of learning two languages on metalinguistic awareness. *Cognition*, 34, 1–56.

Gardner-Chloros, P. (1995). Codeswitching in community, regional and national repertoires: the myth of the discreteness of linguistic systems. In L. Milroy and P. Muysken (eds), *One Speaker, Two Languages* (pp. 68–89). Cambridge: Cambridge University Press.

Genesee, F. 2002. Portrait of the bilingual child. In V. Cook (ed.), *Portraits of the L2 User* (pp. 161–79). Clevedon: Multilingual Matters.

Genesee, F., Nicoladis, E., and Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language*, 22, 611–31.

Goss, N., Ying-Hua, Z., and Lantolf, J. P. (1994). Two heads may be better than one: mental activity in second language grammaticality judgements. In Tarone, Gass, and Cohen 1994, pp. 263–86.

Grondin, N. and White, L. (1996). Functional categories in child L2 acquisition of French. *Language Acquisition*, 5, 1–34.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36, 3–15.

Halliday, M. A. K., McIntosh, A., and Strevens, P. (1964). *The Linguistic Sciences and Language Teaching.* London: Longman.

Han, Z. (2004). *Fossilisation in Adult second language acquisition*. Clevedon: Multilingual Matters.

Harrington, M. (1987). Processing transfer: language specific processing strategies as a source of interlanguage variation. *Applied Psycholinguistics*, 8, 351–77.

Haugen, E. (1953). *The Norwegian Language in America.* Philadelphia, PA: University of Pennsylvania Press.

Hawkins, R. and Chen, C. (1997). The partial availability of Universal Grammar in second language acquisition: the "failed functional features" hypothesis. *Second Language Research*, 13(3), 187–226.

Hilles, S. (1991). Access to Universal Grammar in second language acquisition. In L. Eubank (ed.), *Point Counterpoint: Universal Grammar in the Second Language* (pp. 305–38). Amsterdam: Benjamins.

Holm, A. and Dodd, A. (1996). The effect of first written language on the acquisition of English literacy. *Cognition*, 59, 119–47.

Hornby, A. S. (1956). *A Guide to Patterns and Usage in English.* Oxford: Oxford University Press.

Hyltenstam, K. and Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty and M. Long (eds), *The Handbook of second language acquisition* (pp. 539–88). Oxford: Blackwell.

Ingram, D. (1989). *First Language Acquisition.* Cambridge: Cambridge University Press.

Issidorides, D. and Hulstijn, J. (1992). Comprehension of grammatically modified and non-modified sentences by second language learners. *Applied Psycholinguistics*, 13, 147–71.

James, C. (1998). *Errors in Language Learning and Use: Exploring Error Analysis*. London: Longman.

James, C., Scholfield, P., Garrett, P., and Griffiths, Y. (1993). Welsh bilinguals' spelling: an error analysis. *Journal of Multilingual and Multicultural Development*, 14, 4, 287–306.

Jarvis, S. and Pavlenko, A. (eds) (2008). *Crosslinguistic Influence in Language and Cognition*. Abingdon: Routledge.

Kasper, G. (1981). *Pragmatische Aspekte in der Interimsprache.* Tubingen: Narr.

Kasper, G. (1996). The development of pragmatic competence. In E. Kellerman, B. Weltens, and T. Bongaerts (eds), *EUROSLA 6: A Selection of Papers*, 55, 2, 103–20.

Kasper, G. and Kellerman, E. (eds) (1997). *Communication Strategies: Psycholinguistic and Sociolinguistic Perspectives*. London: Longman.

Keenan, E. L. and Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8, 63–99.

Kellerman, E. and Sharwood-Smith, M. (eds) (1986). *Crosslinguistic Influence and Second Language Acquisition*. Oxford: Pergamon.

Khanna, A. L., Verma, M. K., Agnihotri, R. K., and Sinha, S. K. (1998). *Adult ESOL Learners in Britain.* Clevedon: Multilingual Matters.

Kilborn, K. and Cooreman, A. (1987). Sentence interpretation strategies in adult Dutch-English bilinguals. *Applied Psycholinguistics*, 8, 415–31.

Kirsner, K. (1986). Lexical function: is a bilingual account necessary. In J. Vaid (ed.), *Language Processing in Bilinguals: Psycholinguistic and Neurolinguistic Perspectives* (pp. 21–45). Mahwah, NJ: Erlbaum.

Klein, W. and Perdue, C. (1992). *Utterance Structure: Developing Grammars Again.* Amsterdam: Benjamins.

Klein, W. and Perdue, C. (1997). The basic variety (or: couldn't natural languages be much simpler? *Second Language Research*, 13(4), 301–47.

Krashen, S., Scarcella, R., and Long, M. (eds) (1982). *Child–Adult Differences in Second Language Acquisition.* New York: Newbury House.

Kroll, J. F. (1993). Accessing conceptual representation for words in a second language. In R. Schreuder and B. Weltens (eds), *The Bilingual Lexicon*. Amsterdam: Benjamins.

Kroll, J. and De Groot, A. (eds) (2005). *Handbook of Bilingualism*. Oxford: Oxford University Press.

Lado, R. (1957). *Linguistics across Cultures.* Ann Arbor, MI: University of Michigan Press.

Lakshmanan, U. (1991). Morphological uniformity and null subjects in child second language acquisition. In L. Eubank (ed.), *Point Counterpoint: Universal Grammar in the Second Language* (pp. 389–410). Amsterdam: Benjamins.

Lambert, W. E. (1990). Persistent issues in bilingualism. In B. Harley, P. Allen, J. Cummins, and M. Swain (eds), *The Development of Second Language Proficiency.* Cambridge: Cambridge University Press.

Lantolf, J. (ed.) (2000). *Sociocultural Theory and Second Language Learning.* Oxford: Oxford University Press.

Lenneberg, E. (1967). *Biological Foundations of Language.* New York: Wiley & Sons.

Leopold, W. (1947). *Speech Development of a Bilingual Child: A Linguist's Record. Volume II: Sound Learning in the First Two Years.* Evanston, IL: North-Western University Press.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

Liceras, J. M. (1989). On some properties of the "pro-drop" parameter: looking for missing subjects in non-native Spanish. In S. Gass and J. Schachter (eds), *Linguistic Perspectives on Second Language Acquisition* (pp. 109–33). Cambridge: Cambridge University Press.

Lieberman, P. (1984). *The Biology and Evolution of Language.* Cambridge, MA: Harvard University Press.

Lightbown, P. (1987). Classroom language as input to second language acquisition. In C. W. Pfaff (ed.), *First and Second Language Acquisition Processes.* New York: Newbury House.

Locastro, V. (1987). Aizuchi: a Japanese conversational routine. In L. E. Smith (ed.), *Discourse across Cultures* (pp. 101–13). New York: Prentice Hall.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie and T. K. Bhatia (eds), *Handbook of Second Language acquisition* (pp. 413–68). San Diego, CA: Academic Press.

Lyster, R. and Ranta, L. (1997). Corrective feedback and learner uptake: negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19, 1, 37–66.

Macaro, E. (2005). Code-switching in the L2 classroom: a communication and learning strategy. In E. Llurda (ed.) *Non-Native Language Teachers: Perceptions, Challenges and Contributions to the Profession* (pp. 63–84). Boston: MA, Springer.

Macaro, E. (2006). Strategies for language learning and for language use: revising the theoretical framework. *Modern Language Journal*, 90, 320–37.

MacWhinney, B. (1987). Applying the competition model to bilingualism. *Applied Psycholinguistics*, 8, 315–27.

Major, R. C. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology.* Mahwah, NJ: Erlbaum.

Major, R. C. (2002). The phonology of the L2 user. In V. Cook (ed.), *Portraits of the L2 User* (pp. 65–92). Clevedon: Multilingual Matters.

Makino, T. (1993). *Perspectives on Second Language Acquisition.* Tokyo: Yumi Press.

McDonough, S. (1995). *Strategy and Skill in Learning a Foreign Language.* London: Edward Arnold.

Meisel, J. M. (1990). Early differentiation of languages in bilingual children. In K. Hyltenstam and L. K. Obler (eds), *Bilingualism across the Lifespan* (pp. 13–40). Cambridge: Cambridge University Press.

Meisel, J. M., Clahsen, H., and Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition*, 3, 2, 109–35.

Meisel, J. M. and Muller, N. (1992). Finiteness and verb placement in early child grammars. In J. M. Meisel (ed.), *The Acquisition of Verb Placement* (pp. 109–38). Dordrecht: Kluwer.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32, 543–63.

Morgan, J. (1986). *From Simple Input to Complex Grammar.* Cambridge, MA: MIT Press.

Muysken, P. (1995). Codeswitching and grammatical theory. In L. Milroy and P. Muysken (eds), *One Speaker, Two Languages* (pp. 177–98). Cambridge: Cambridge University Press.

Myers-Scotton, C. M. (2005). Supporting a differential access hypothesis; code-switching and other contact data. In J. Kroll and A. De Groot (eds), *Handbook of Bilingualism* (pp. 326–48). Oxford: Oxford University Press.

Naiman, N., Fröhlich, M., Stern, H., and Todesco, A. (1995). *The Good Language Learner.* Clevedon: Multilingual Matters.

Nation, I. S. P. (2001). *Learning Vocabulary in Another language.* Cambridge: Cambridge University Press.

Nemser, W. (1971). Approximative systems of foreign language learner. *International Review of Applied Linguistics*, 9, 115–24.

Newport, E. L. (1976). Motherese: the speech of mothers to young children. In N. Castellan, D. Pisoni, and G. Potts (eds), *Cognitive Theory: Vol. 2.* Hillsdale, NJ: Erlbaum.

O'Grady, W. (2005). *How Children Learn Language.* Cambridge: Cambridge University Press.

O'Malley, J. and Chamot, A. (1989). *Learning Strategies in Second Language Acquisition.* Cambridge: Cambridge University Press.

Oksaar, E. (1970). Zum Spracherwerb des Kindes in Zweisprachiger Umgebung. *Folia Linguistica*, 4, 330–58.

Otsu, Y. and Naoi, K. (1986). Structure-dependence in L2 acquisition. Paper presented at JACET Keio University, Tokyo, September.

Oyama, S. (1976). A sensitive period in the acquisition of a non-native phonological system. *Journal of Psycholinguistic Research*, 5, 261–85.

Paap, K. R., Noel, R. W., and Johansen, L. S. (1992). Dual-route models of print to sound: red herrings and real horses. In R. Frost and L. Katz (eds), *Orthography, Phonology, Morphology, and Meaning* (pp. 293–318). Amsterdam: Elsevier.

Page, N. (1986). *A Conrad Companion.* Basingstoke: Macmillan.

Paivio, A. and Desrochers, A. (1979). Effects of an imagery mnemonic on second language recall. *Canadian Journal of Psychology*, 33, 17–28.

Palmer, H. E. (1926). *The Principles of Language Study.* London: Harrap.

Paradis, J. and Genesee, F. (1996). Syntactic acquisition in bilingual children: autonomous or independent? *Studies in Second Language Acquisition*, 18, 1–25.

Patkowski, M. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, 30, 449–72.

Pica, T., Young, R., and Doughty, C. (1987). The impact of interaction on comprehension. *TESOL Quarterly*, 21, 737–58.

Pienemann, M. (1998). *Language Processing and Second-Language Development: Processability Theory.* Amsterdam: John Benjamins.

Pienemann, M. and Johnston, M. (1987). Factors influencing the development of language proficiency. In D. Nunan (ed.) *Applying Second Language Acquisition Research.* Adelaide: National Curriculum Resource Centre.

Piller, I. (2002). *Bilingual Couples Talk: The Discursive Construction of Hybridity.* Amsterdam: John Benjamins.

Poplack, S. (1980). Sometimes I'll start a sentence in English y termino en espanol. *Linguistics*, 18, 581–616.

Poulisse, N. (1990). *The Use of Compensatory Strategies by Dutch Learners of English.* The Hague: Mouton.

Poulisse, N. (1996). Strategies. In P. Jordens and J. Lalleman (eds), *Investigating Second Language Acquisition* (pp. 135–63). The Hague: Mouton.

Pulvermüller, F. and Schumann, J. (1994). Neurobiological mechanisms of language acquisition. *Language Learning*, 44, 681–734.

Queen, R. M. (2001). Bilingual intonation patterns: evidence of language change from Turkish-German bilingual children. *Language in Society*, 30, 55–80.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1972). *A Grammar of Contemporary English.* Longman: London.

Radford, A. (1990). *Syntactic Theory and the Acquisition of English Syntax.* Oxford: Blackwell.

Regan, V. (1998). *Contemporary Approaches to Second Language Acquisition in a Social Context.* Dublin: University College Dublin Press.

Rizzi, L. (2004). On the study of the language faculty: results, developments and perspectives. *Linguistic Review*, 21, 323–44.

Romaine, S. (1994). *Bilingualism.* Oxford: Blackwell.

Roulet, E. (1973). *Linguistic Theory, Language Description and Language Teaching.* London: Longman.

Schachter, J. (1988). Second language acquisition and its relationship to Universal Grammar. *Applied Linguistics*, 9, 3, 219–35.

Schnitzer, M. L. and Krasinski, E. (1997). The development of segmental phonological production in a bilingual child. *Journal of Child Language*, 21, 585–622.

Schwartz, B. and Sprouse, R. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12, 1, 40–72.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 3, 209–31.

Selinker, L. (1992). *Rediscovering Interlanguage.* London: Longman.

Sharwood-Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, 15, 165–79.

Singleton, D. and Ryan, L. (2004). *Language Acquisition: The Age Factor*, 2nd edn. Clevedon: Multilingual Matters.

Snow, C. and Hoefnagel-Höhle, M. (1978). The critical age for language acquisition: evidence from second language learning. *Child Development*, 49, 1114–28.

Sridhar, K. K. and Sridhar, S. N. (1986). Bridging the paradigm gap; second language aquisition theory and indigenised varieties of English. *World Englishes*, 5, 1, 3–14.

Stocker-Edel, A. (1977). The responses of Wellingtonians to a foreigner's English. *Archivum Linguisticum*, 8, 13–27.

Swain, M. and Wesche, M. (1975). Linguistic interaction: case study of a bilingual child. *Language Sciences*, 37, 17–22.

Taeschner, T. (1983). *The Sun Is Feminine.* Berlin: Springer.

Tarone, E. E. (1980). Communication strategies, foreigner talk, and repair in interlanguage. *Language Learning*, 30, 2, 417–31.

Tarone, E. E., Gass, S. M., and Cohen, A. D. (eds) (1994). *Research Methodology in Second Language Acquisition.* Mahwah, NJ: Erlbaum.

Tomasello, M. (ed.) (1998). *The New Psychology: Cognitive and Functional Approaches to Language Structure.* New York: Psychology Press.

Toncheva, E. (1988). Errors in the use of articles in the interlanguage of Bulgarian learners of English. In A. Danchev (ed.), *Error Analysis – Bulgarian Learners of English* (pp. 127–9). Sofia: Narodna.

Towell, R. and Hawkins, R. (1994). *Approaches to Second Language Acquisition.* Clevedon: Multilingual Matters.

Trahey, M. and White, L. (1993). Positive evidence and pre-emption in the second language classroom. *Studies in Second Language Acquisition*, 15, 181–204.

Vainikka, A. and Young-Scholten, M. (1991). Verb raising in second language acquisition: the early stages. *Theories des Lexikons*, 4, Dusseldorf University.

Vainikka, A. and Young-Scholten, M. (1996). Gradual development of L2 phrase structure. *Second Language Research*, 12, 1, 7–39.

Volterra, V. and Taeschner, R. (1978). The acquisition and development of language by bilingual children. *Journal of Child Language*, 5, 311–26.

Vygotsky, L. S. (1962 [1934]). *Thought and Language.* Cambridge, MA: MIT Press.

Vygotsky, L. S. (1978 [1935]). *Mind in Society.* Cambridge, MA: Harvard University Press.

Watson, I. (1991). Phonological processing in two languages. In E. Bialystok (ed.), *Language Processes in Bilingual Children* (pp. 25–48). Cambridge: Cambridge University Press.

Weinreich, U. (1953). *Languages in Contact.* The Hague: Mouton.

White, L. (1986). Implications of parametric variation for adult second language acquisition: an investigation of the pro-drop parameter. In V. J. Cook (ed.), *Experimental Approaches to Second Language Acquisition* (pp. 55–72). Oxford: Pergamon.

White, L. (1991). The verb-movement parameter in second language acquisition. *Language Acquisition*, 1, 337–60.

White, L. and Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research*, 17, 1.

Wieden, W. and Nemser, W. (1991). *The Pronunciation of English in Austria.* Tübingen: Gunter Narr.

Williams, L. (1977). The perception of consonant voicing by Spanish English bilinguals. *Perception and Psychology*, 21, 4, 289–97.

Wode, H. (1981). *Learning a Second Language.* Tübingen: Narr.

Wolfe Quintero, K. (1992). Learnability and the acquisition of extraction in relative clauses and wh-questions. *Studies in Second Language Acquisition*, 14, 1, 39–71.

Zandvoort, R. W. (1957). *A Handbook of English Grammar.* London: Longman.

Part VI Applications of Linguistics

28 Second Language Pedagogy

Where Theory Meets Practice

JANIE REES-MILLER

Foreign language or second language (L2) pedagogy is where theory meets practice. The L2 teacher must answer crucial questions of theory: What is the nature of language? How are L2s acquired? The L2 teacher must also answer general practical questions: Who are my students? Why are they learning this language? Then, given who my students are and their reasons for learning the language, what are reasonable goals and expectations for my students? The answers to these general questions of theory and practice will crucially inform the day-to-day questions of practice: How are we going to spend our time in class? What materials will we use? What activities or tasks will we do?

This chapter will describe the relationship of theory and practice in various language teaching methodologies with particular emphasis on communicative language teaching, which has become the most commonly used approach to L2 pedagogy today. In light of recent developments, the chapter will reconsider the theory/practice relationship in the second decade of the twenty-first century and examine the challenges posed to our underlying theories by the emergence of English as a worldwide lingua franca.

1 Methodologies[1]

In the last fifty to sixty years, the questions posed in the first paragraph about theory and practice have been answered in quite different ways, as the following examples will serve to demonstrate. As the first example, the book *Practical French Review* (Barton and Sirich 1954, first published 1941) organizes the table of contents according to grammar topics. Lesson 1 is devoted to articles and prepositions with names of countries and cities. The first four pages of the lesson are taken up with grammar explanation in English of the topics listed in the contents, followed by a reading passage concerning Pierre and his late appearance at lunch due to delays at the American Express. This is followed by comprehension questions in French and a "composition" written in English for translation to French, grammar exercises to be translated from English to French, and an oral drill: "Translate and reply with a complete sentence: 1. When are you going to be in French Indo-China? 2. What languages does Dr. Renauld speak? 3. What time do you have French class on Mondays? . . ." Aside from the reading and composition, the grammar exercises

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

are sentence-level tasks unrelated to each other or to the lives of the students. Although the authors of the book do not have a preface setting forth their theoretical bent, it is clear that the core of language is conceived to be discrete language points. We can infer from the exercises that languages are assumed to be learned through reading and translation. Since the reading and translation passages concern an American visiting France, we might infer that a possible goal for students is a sojourn in France. We can envision a classroom in which there is teacher-fronted explanation of grammar in English, perhaps some reading aloud of the passages, assigned homework, and some teacher-led sentence-level work.

A contrast to *Practical French* is *Eastern Arabic*, a course in colloquial Arabic (Rice and Sa'id 1966). The table of contents is organized according to situations and language functions such as greetings, directions, telling time, in a coffee shop, and buying fruits and vegetables. The first lesson begins with "pattern sentences," essentially a dialogue of greetings and introductions. Following the dialogue are "structure sentences" with examples of feminine forms, plural forms, and independent pronouns. There is no grammar explanation accompanying the examples, just a list of expressions and their translation, e.g., *How are you? Please sit down. Will you excuse me? I'm well. He's well. She's well*. A section of grammatical explanations follows and includes the topics of word stress and prominence, sentences without verbs, and independent pronouns. The lesson ends with drills; there are sentences in Arabic that would be said to a man that must be changed to sentences that would be said to a woman and to a group. This book does have a fairly extensive preface that sets forth certain principles for use. It states that the book is designed to teach spoken language and is "based on the principle that the only way to learn to understand a spoken language is to hear it spoken, and the only way to learn to speak it is to practice speaking." The teacher is to be a native speaker of the language, whose pronunciation of the dialogues is to be mimicked; dialogues must be repeated after the teacher's model, and then memorized. The instructor is directed to "help [the student] form proper habits, and try to keep him from continuing wrong ones." Although the dialogues do not derive from students' personal experience, the book was published in Beirut (albeit first compiled at Georgetown University) and used for language instruction in Arabic classes for foreigners in-country; topics throughout the book relate to common situations that a learner of Arabic would face in the Middle East.

A third book, in contrast to the previous two, is *Claro Que Sí!* (Garner, Rusch, and Domínguez 2008), an introductory Spanish book for US university students. The table of contents is headed in Spanish, and each chapter contains one or more sections for listening, essential vocabulary, reading, culture, and communicative grammar (*gramática para la comunicación*, e.g., talking about yourself and others, expressing likes and dislikes, expressing habitual and future actions, etc.). Like *Eastern Arabic*, the first lesson begins with a dialogue of people introducing themselves and saying where they are from. The dialogue, though, is presented via a video of two South American students in Bolivia. Following this are activities in which students in the class mingle with other class members and introduce themselves. There are a dizzying number of separate activities and bits of information, including various forms of greetings and information about hand-shaking and cheek-kissing, information on use of *tu* and *usted* (the formal and informal "you"), and map work on countries and capitals in the Spanish-speaking world. The concept of language here is multifaceted: language includes grammar and vocabulary to be sure, but also functions as communication within the social realm, which is inextricably bound up in the culture(s) in which the language is used. In the preface, students are told that "the ultimate goal is communication" and are urged to participate actively in oral activities. While listening, reading, or viewing, students are cautioned to focus on gist without expecting to understand every word, and to view studying Spanish as an adventure in which they should be willing to make mistakes.

These examples from three different texts provide very different answers to the questions posed at the beginning of the chapter and exemplify modified forms of three methodologies or approaches used in L2 pedagogy in the last half century, namely grammar translation, audiolingualism, and communicative language teaching (CLT). Each had a very different starting point

because of differing theoretical concepts of what language is and how languages are learned. Grammar translation, illustrated by *Practical French*, took as a starting point language as discrete grammatical points and vocabulary items that are best learned through the medium of the first language. The audiolingual approach, illustrated by *Eastern Arabic*, took spoken language as primary, and the model of spoken language must be the native speaker. It presupposed that language learning was a process of habit formation best accomplished through imitation and repetition to eliminate the possibility of error. *Claro Que Sí!* illustrates communicative language teaching (CLT), which has been the dominant theoretical model since the last quarter of the twentieth century. In CLT, language is conceptualized as not just a framework within the learner's head but a communicative system that relies on use within a social context. Thus, language includes not just grammar rules or correct pronunciation but the way in which individuals interact effectively with each other.

2 Communicative Language Teaching (CLT)

Since CLT is far and away the dominant model for L2 pedagogy worldwide, a more detailed description of what that entails is appropriate. CLT had its origin in the late 1960s and early 1970s, both in Europe and the USA. As Europeans considered the language needs of a more unified Europe, it became clear that all Europeans needed to be multilingual (Trim 1979 [1973]). Also, as educational opportunities became available in English for an increasing number of citizens in the world, it became apparent that previous concepts of language and modes of teaching were inadequate (Widdowson 1979 [1972]).

2.1 CLT: theoretical bases and implications for practice

The theory of language underlying CLT derives from work in the 1960s and 1970s that focused on language as communication rather than just a grammatical system in an individual's mind. In an oft-cited essay derived from a talk, Dell Hymes (1979 [1972, 1966]) took issue with Noam Chomsky's definition of linguistic competence as that of an ideal native speaker in a homogeneous monolingual community. Hymes pointed out that the ideal native speaker does not exist, speech communities are not homogeneous, and language cannot be divorced from its sociocultural uses for communication. He proposed the term *communicative competence*, which takes account not only of whether something is formally possible but whether it is appropriate within a given social context.

Hymes' idea converged with ideas that had been developed in Europe; these ideas had strongly influenced the functional/notional syllabus proposed by the Council of Europe (van Ek 1979 [1975]) for foreign language education of Europeans. Functions refer to how language is used to accomplish certain social actions, such as asking permission, expressing likes and dislikes, apologizing, requesting, and so on. Notions refer to semantic concepts that can be expressed through language, such as past vs. present, existence vs. nonexistence.

Meanwhile, in the USA in the early 1970s, Savignon was developing classroom activities that would reflect this new concept of language as communicative. Since language is conceived as essentially social, it follows that students would learn language through interaction. Learners would be able to take risks with the language and make mistakes as they sought to make meaning together by interacting with each other (Savignon 2001). Activities used in a CLT classroom should then involve the students in tasks that have a communicative purpose; students should be using the target language for authentic purposes, in other words, for communicating real information about themselves and their world. Similarly, CLT makes use wherever possible of authentic texts, texts that are not artificially modified or specially written for non-native speakers but reflect real language as it is used by native speakers.

As the theoretical bases of CLT have been elaborated over the last decades, they have drawn on work in second language acquisition. One of the most influential theoretical ideas is that of comprehensible input, which refers to input that is slightly beyond the learners' existing competence but is nonetheless understandable (Krashen and Terrell 1988: 32–7). The practical application of this idea is that, in the CLT classroom, the teacher is expected to use the target language almost exclusively and thus provide comprehensible input. But the focus is not just on the teacher's input; students can get comprehensible input from interacting with each other. Negotiation of meaning refers to the way in which interlocutors cooperate to make meaning together by asking for repetition or clarification and checking and confirming comprehension. Through negotiation of meaning, learners hone their strategic competence and also obtain comprehensible input from their partners (Pica et al. 1996; also discussed in Macaro 2003: 186–7). Output is also important not only to help learners practice what they have learned but also to push them to stretch their language resources to make themselves understood (Swain 1985; also discussed in Gass and Selinker 2001: 276–8). Thus, CLT lessons include interactive classroom activities to encourage both negotiation of meaning and output.

Because learners are expected to engage in frequent interactive activities in a CLT classroom, the teacher takes on a multitude of roles, only one of which may be the language authority. Far more important may be the role of organizer of language activities and facilitator for language tasks. The communicative task is a kind of experiential learning activity in which learners must use the target language resources at their disposal to arrive at an outcome (Richards and Renandya 2002: 94; Norris 2011: 579–80). Such tasks can vary from conversations exchanging personal information and opinions to role-play or information gap activities, in which students must work collaboratively in pairs or small groups and share information to complete a task successfully. These kinds of activities are underlain by the assumptions about the nature of language as social, communicative, and authentic, and by assumptions about second language acquisition as requiring negotiation of meaning and social engagement.

The relationship of the four skills of speaking, listening, reading, and writing in CLT contrasts with earlier methods. In grammar translation, reading and writing were the more important skills while speaking and listening got short shrift. In audiolingualism, because speaking was considered primary, students were taught speaking skills in pronunciation drills and memorized dialogues first; in some cases of audiolingual application, students were not even supposed to see the printed form of the language until they had acquired good habits in speaking (see, e.g., Brooks 1964: 51–2; Lado 1964: 61–9).

Generally, for CLT, the four skills are integrated, with special emphasis given to speaking and listening since the theoretical bases of the approach describe language as communication, which presupposes speaking and listening as part of the communicative process. However, since reading and writing also play a role as part of the communicative process, reading and writing may be integrated into work on speaking and listening. A "typical" CLT lesson might include a listening comprehension passage and a related speaking activity. Some kind of realistic or authentic reading activity may also be included with a follow-up writing task. For example, if the theme is ordering food in a restaurant, students might be presented with an actual restaurant menu and listen to a recording in which various diners place orders for food. This could be followed by a role-play activity in which students practice placing orders from the authentic menu. Combining skills of both speaking and writing, students might then be asked to work in groups to create a menu for a new restaurant. Depending on the particular theoretical inclinations of the instructor or textbook, there may be a focus on grammatical forms such as (for English) the distinction between *I'll have . . .* and *I'm going to have . . . .* Alternatively, there could be a more functional approach with a focus on a variety of grammatically unrelated ways of placing an order, such as *A bowl of tomato soup, please*; *Tomato soup for me; I'd like the tomato soup*; *Could I have the tomato soup?*

2.2 CLT and the question of grammar

Virtually since the inception of CLT, theorists and practitioners alike have vigorously debated the role of grammar instruction. Should grammar be explicitly taught, or will students acquire it as they go along? If grammar is taught, in what ways should it be taught? Should it be a stand-alone part of the course, or should it be integrated into whatever content the language is being used to discuss? A related question has to do with error correction. Assuming that students will inevitably make mistakes, as doing so is conceived to be part of the process of learning, how should errors be treated? Should errors be corrected explicitly, and at what stage of the process should errors be corrected?

There are those within the CLT tradition on either extreme of the grammar instruction question. On the one hand, Krashen and Terrell's (1988) Natural Approach distinguishes *learning*, which they view as superficial knowledge *about* the target language, from *acquisition*, which they view as the deeper knowledge *how to use* the target language. They hypothesize that comprehensible input should be sufficient for acquisition. In a classroom that puts into practice the ideas about grammar embodied in the Natural Approach, there would be little if any explicit grammar instruction, and error correction might be through recasts in which the teacher simply repeats what the student has said in the correct form.

On the other extreme is the structural syllabus, which is organized according to pre-set grammar points. It is supported by the so-called PPP procedure for teaching new language. PPP – standing for *p*resentation, *p*ractice, *p*roduction – is still in common use today (Swan 2005; Larsen-Freeman 2011: 523–4) and was the lesson plan style favored by teacher training programs of the Royal Society of Arts in the 1980s (see, e.g., Matthews, Spratt, and Dangerfield 1985: 5–17 or Gower and Walters 1983: 65–142; the Royal Society of Arts programs are now part of Cambridge ESOL). In PPP, new grammar or vocabulary structures are first presented in some kind of context, perhaps in a reading or listening text. The teacher then highlights the form to be learned and provides students with an opportunity for controlled or semi-controlled practice of the form.

During the controlled or semi-controlled practice phase, the teacher provides error correction to assure that students produce the form correctly. In both presentation and practice of form, contextualization has long been an important feature of teaching grammar within the communicative approach (Dickens and Woods 1988). (This contrasts sharply with decontextualized audiolingual pattern drills, such as examples given by Lado (1964: 99): "The students are busy → Are the students busy? The teacher is reading papers. → Is the teacher reading papers? The class can go faster. → Can the class go faster? . . .") In the production phase of the PPP, the teacher sets up an open-ended communicative task in which learners can produce the form creatively. During the production phase, the teacher monitors students as they work and provides feedback on errors when the task is completed.

A middle ground seems to be claimed by scholars who eschew a syllabus based on a preset list of grammar points but support some sort of form-focused instruction as a complement to communicative activities (e.g., Long 2011). Spada and Lightbown (2008) see a role for both integrated and isolated form-focused instruction within CLT. If learners are engaged in a communicative task and the teacher intervenes briefly with correction or explanation to help them express meaning more accurately, the form-focused instruction is integrated within the activity. Or it may be isolated from the communicative task as instruction either to prepare the learners with language forms they will need for doing the task or, after task completion, to correct language the learners produced.

Despite the voluminous literature published on the subject, both teachers and students expect grammar instruction in some form to be part of a language course. In response to consumer demand, publishers continue to produce language textbooks with grammar instruction. Through an examination of some of the best-selling English language textbooks, Waters (2012) demonstrates that a focus on grammar has remained a constant element in course books and concludes

that "enduring situational realities" will continue to drive classroom practice, rather than "some kind of theory-driven 'second coming'" (p. 448).

2.3 CLT applications: content-based instruction (CBI) and task-based instruction (TBI)

Two developments within the CLT tradition warrant mention as they have become quite extensively used worldwide for teaching a variety of languages. These two developments are content-based instruction (CBI) and task-based instruction (TBI; also known as TBLT for task-based language teaching). CBI and TBI are related to each other in the sense that both take as a starting point the idea that language is best learned when the language itself is the tool for *doing* something rather than the object of learning. As defined by Stryker and Leaver (1997), content-based instruction refers to the integration of language learning with content. In fact, the content to be learned, be it culture, literature, journalism, or banking, is the focus of the course; the second or foreign language is the means by which the content is learned. In addition to having the subject matter as the core, authentic texts and language are used. (With reference to texts, *authentic texts* are produced by native speakers for native speakers; *authentic language* refers to language that would be used to accomplish real-world goals, although cf. Badger and McDonald (2010) on questions concerning the meaning of *authentic*.) A third characteristic of CBI is the tailoring of the course to meet the specific needs of a particular group of students.

Content-based instruction (CBI) has been used for K-12 immersion education in such diverse settings as the USSR in the 1960s–1980s, Canadian French immersion schools from the 1960s to the present, and US Spanish–English bilingual education programs at the present. In these school settings, children are taught the range of normal school subjects in the medium of the L2. At the level of tertiary and adult education, CBI has been utilized in US universities, for example, for teaching business in French or anthropology in Japanese (Stryker and Leaver 1997: 15–21) and more recently with various models of languages across the curriculum (Klee 2009). Since late 2000, a few Language Flagship Programs have been developed in the US, using a combination of CBI, study abroad, and internships for intensive language training of undergraduate students in critical languages (Spring 2012).

The European Union has been especially active in promoting CBI, which is called content and language integrated learning, or CLIL. Examples of CLIL that have received funding from the EU include secondary school science courses in physics, chemistry, biology, and technology taught in English in Sweden; courses taught in English in a nursing school in Grenoble to enable French nurses to admit patients, apply therapeutic measures, and deal with health-related topics in English; and business ethics taught in French to final-year undergraduates majoring in fields related to international business at Dublin City University in Ireland (European Commission 2006). CLIL seems to have many positive outcomes (e.g., Lorenzo, Casal, and Moore 2010); however, conclusions about the degree of success may need to be tempered by close attention to whether CLIL groups are truly comparable to non-CLIL groups with whom their results have been compared (Dalton-Puffer 2011; also Bruton 2011).

Similar to CBI, task-based instruction (TBI) rests on the belief that the purpose of learning language is language use, and that learning takes place by making use of language. A distinction can be made between the "weak form" of TBI, which makes use of communicative tasks simply as part of CLT, and the "strong form" of TBI, which organizes the whole syllabus around tasks (McDonough and Caikitmongkol 2007). The strong form of TBI was pioneered in the Bangalore Project in India and publicized by Prabhu (1987) and has since been used extensively in Flanders to teach Dutch as a second language (Van den Branden 2006: 13–14). TBI is gaining in popularity in Asia as well: Hong Kong included TBI in its syllabus in the late 1990s, and China mandated use of communicative tasks in 2001 (Butler 2011). According to the proponents of TBI, language is not

learned in isolated bits but rather in a complex mapping of form to content. In the strong form of TBI, the syllabus is organized around a series of tasks, goal-directed activities that require language for successful performance (Van den Branden 2006: 3–6; Ellis 2009: 223). Specifically, a task is characterized as having a specific discernible outcome with real-world relevance. Performance of the task should be focused on *meaning*, not language form, and can call on any one or more of the skills of listening, speaking, reading, and writing (Willis 2004: 13–16; Ellis 2009: 223–4). The task cycle begins with pre-task material introduced by the teacher; this could include introduction to the topic and an explanation of the upcoming task. The students are then set to work on the task. On completion of the task, they plan their report on the task and subsequently present the results of their task to the class. Following the task completion and report is the language focus part of the task cycle. At this time, the students are encouraged to review and analyze the language features of both the input they received and the output they produced. Finally, the teacher provides practice with the new language forms that were required for successful completion of the task (Willis 2004: 42). Other advocates of TBI allow for attention to language form at various stages of the task cycle, not restricting it solely to the post-task phase (Ellis 2009: 231–2).

2.4 *Criticisms of CLT, CBI, and TBI*

Despite its widespread use, CLT is not without its critics, particularly when it involves teaching a *foreign* language, one that is not used in the environment in which it is taught (e.g., German or French in most parts of the USA or English in Vietnam or China). Any methodology betrays a certain value system, as discussed in some detail by Adamson (2006: 610–11, 615), who states, "methods only cross cultural boundaries easily if they can be appropriated in a form consistent with the values and beliefs of the community adopting and adapting the methods." With its emphasis on student-centered teaching, CLT embodies values and expectations of individualism and creativity on the part of the students that may be at odds with prevailing educational traditions and culture (e.g., see Cortazzi and Jin 1996; Shamim 1996; Butler 2011 offers a more nuanced summary).

Local teaching conditions will also have an impact on how successfully any methodology can be employed. In the case of CLT, students are expected to use the target language among themselves in the classroom, which is a reasonable expectation in a multilingual language center (e.g., for immigrants in the USA or in Europe). However, when the students share the same first language and do not need the L2 outside the classroom, there may be little incentive for them to participate in pair or group work in the target language. In addition to these difficulties, Hiep (2007) and Butler (2011) also note that practices used in a second language setting in the West cannot necessarily be imported wholesale into a country such as Vietnam or China because of different sociocultural, political, and physical conditions. Large classes, differing concepts of students' and teachers' appropriate roles, concern with passing national grammar-based exams, and teachers who may not be proficient enough in the target language to teach with confidence through the target language can all impede the successful implementation of CLT (Butler 2011; Li 1998).

Some of the concerns voiced by Butler (2011) and by Hiep (2007) apply wherever CLT is used for foreign language teaching. In the USA, Schulz (2006) complains that the goal of CLT of communicative competence is "neither a realistic nor a sufficient goal" for US university students who typically take only the first two years of a foreign language to satisfy university general education requirements. Magnan (2007) notes that personalized CLT activities as used in foreign language classes in the USA lack cultural authenticity. When US students talk about themselves with each other, albeit in the target language, their ideas about language and culture are bounded by their own experiences and merely serve to reinforce US-centered views. In such activities, US students are not getting the experience of language as "socially constituted" (Magnan 2007: 250).

Similar practical concerns limit the effectiveness of content-based instruction. The success of CBI rests on practical considerations that are often far removed from specifically linguistic concerns. Of crucial importance is the integration of both content and language. If comprehension of content is stressed, with language instruction relegated to a subordinate role, then students' language skills may not improve (Klee 2009; Tan 2011). If, on the other hand, language instruction takes priority, then content is watered down (Lyster and Ballinger 2011). In their introduction to a special issue devoted to CBI in the journal *Language Teaching Research*, Lyster and Ballinger (2011) stress the crucial role of professional development for both language and content teachers. However, teacher training and the development of successful CBI models at all levels are expensive and time consuming; when funding dries up, so do the CBI programs (Klee 2009).

Task-based instruction (TBI) in particular has come in for its share of criticism based on both theory and implementation (although see Ellis 2009 for refutation of some criticisms). Critics point to flaws in the underlying theory, which rests on the assumptions that comprehensible input is sufficient for acquisition (Sheen 1994) and that language learning is taking place while students are engaged in communicative tasks (Swan 2005). Proponents of TBI are accused of ignoring research that would contradict TBI practice. Sheen (1994), for example, cites research results that show the effectiveness of explicit, deductive grammar instruction, while Swan (2005) points to skill-building research that indicates the importance of practice to turn declarative knowledge into automatized procedural knowledge. When students are supposed to be acquiring new language, it is through interacting with each other, and Swan (2005) questions whether they are receiving sufficient quantity and frequency of exposure to new language. In practical terms, teachers need knowledge, training, and commitment to the principles in order to implement TBI effectively. If TBI is simply imposed from above by government mandate, teachers may not understand the purpose, discipline may be problematic, and students may not actually be using the L2 (Carless 2004). As Butler (2011: 48) points out, the more engaging the task, the more likely students are to use their L1 in order to complete the task efficiently, and that then calls into question the value of the tasks for L2 learning. For reasons such as these, critics consider the exclusive use of TBI in a foreign language environment to be inadequate. That being said, some of the criticisms of implementation can be offset if teachers themselves create the program to fit their local needs and then pilot, monitor, and revise the program in response to input from teachers and students (McDonough and Chaikitmongkol 2007).

3 The Postmethods Era

The heyday of methods, according to Richards and Rogers (2001: 14–16), was the 1950s to 1980s, when correct use of a particular method or approach was thought to yield better results than using another method. Indeed, this coincided with a time when education in general was preoccupied with finding the best method (e.g., New Math of the 1960s or Whole Language of the 1980s). The idea was that the best method would improve the quality of instruction if teachers just adhered to the classroom techniques and activities associated with it.

However, teachers can become cynical and are often frustrated with mandates to switch to a new magic bullet that does not address the complexities of their own particular situation. The almost religious fervor with which some new methods and approaches have been introduced also dismisses the beliefs and previous experience of teachers actually in the classroom (Adamson 2006: 616). The literature of pedagogy has begun now to focus more on the teacher as a mediator of learning, a practitioner who is able to take a principled approach that is pragmatic and reflective (e.g., Richards and Lockhart 1994). Allwright's (2005) Exploratory Practice, for example, encourages practicing teachers and their learners to explore "puzzles" in the classroom with the ultimate goal of achieving deeper understanding of classroom life.

The recognition of the teacher as principled practitioner is long overdue, and some have labeled this trend as the "postmethod condition" (Kumaradivelu 1994, 2001; Brown 2002 [1997]). Kumaradivelu (1994: 29–31) characterized the postmethod condition as: a search for an alternative to the confinement of a single set of practices inherent in adoption of a method; an acceptance of teacher autonomy that allows teachers to respond to their own academic and administrative contexts; and a recognition of principled pragmatism that allows teachers to choose practices that are engaging and effective for their particular students. Brown (2002 [1997]: 10) adds that the demise of methods can be ascribed in part to the lack of distinctiveness of various methods beyond the early stages and an inability to test the effectiveness of methods empirically. Teachers are not cookies cut with the same cutter, and teaching is too "artful and intuitive" to be subjected to study under rigorous experimental conditions.

Another possible contributor to a decreased focus on method and methods, at least in the USA, was the proficiency movement of the 1980s and 1990s. The American Council on the Teaching of Foreign Language (ACTFL) and Educational Testing Services (ETS) combined forces to articulate expected outcomes for each level of foreign language study. The ACTFL standards spell out what learners at a particular level should be able to *do* in the target language (Mitchell and Vidal 2001: 32). For example, in speaking at the general intermediate level, learners should be able to ask and answer questions, creatively combine learned elements, and complete a basic communicative task from start to finish (American Council for the Teaching of Foreign Languages 1983). The standards did not specify *how* goals were to be reached, thus freeing teachers from the grip of adherence to any particular method. Instead, teachers could choose whatever methods or techniques worked for them and their particular students (Mitchell and Vidal 2001: 32).

Bell (2003) links the postmethod condition to postmodernism and a move from positivism to pragmatism, including an awareness of the need to adapt to changing contexts. In fact, as he points out, teachers have never universally followed all the strictures of any given method because an individual teacher's methodology evolves from the interaction of teacher, students, materials, and activities.

Work by an Australian research team (Breen et al. 2001) confirms the individualized nature of teachers' belief systems and reveals teachers to be principled practitioners who make conscious reflective choices in their practice. The research team examined teaching practice and self-identified principles motivating the practice by 18 experienced English as a second language (ESL) teachers. The team found that each individual teacher articulated a coherent set of practices and principles. However, the relationship each individual teacher saw between principle and practice often differed from the relationship posited by other teachers. While the team found a number of shared principles voiced by all the teachers, the practices that teachers identified as arising from the principles varied widely. Similarly, there were common practices, such as pair and group work and teacher modeling and explanation. However, individual teachers attributed these common practices to different principles. The research team concluded that, unlike novices, not only do experienced teachers have more knowledge of subject and craft, but they also structure that knowledge differently as a result of experience in their individual teaching lives.

In her qualitative study of seven experienced ESL teachers in New Zealand, Wette (2009) found that teachers began their pre-course planning with certain very general outcomes in mind and only specified and adapted course content, delivery, and pacing to fit the individual learners in the class once the course was underway. Ur (2013) calls this kind of teaching "situated methodology." Unlike traditionally defined method, which originates in research in applied linguistics, situated methodology is driven by local needs and constraints, including: the characteristics of learners – their age, level, and goals; the demands of stakeholders, such as employers, parents, or government ministries; the content of exams or other high-stakes assessment measures; and the preferences, beliefs, and personality of the individual teacher.

For experienced practicing teachers, it would probably be unrealistic to say that theory is at the forefront of their minds as they plan lessons and teach in the classroom every day. Teachers

are pragmatic and do not necessarily work top-down. It is far more likely that a practicing teacher may try a new technique that appears intriguing, whether she has heard of it at a conference or workshop, read about it in a journal article, or simply had her own bright idea. It is in the aftermath of trying the new technique that the experienced teacher may reflect on why it did or did not work with her students and make her own connections to the principles that guide her teaching.

It must be said here that the success of a particular technique, lesson, or even language course may well boil down to an individual's teacher's skill in the teaching craft, a skill that has little to do with a theory of language or second language acquisition. Instead, it is related to the characteristics of the students in the class, the teacher's knowledge of students' personalities, levels, and learning preferences, and her skill in classroom management. She must be able to choose classroom activities that are both challenging and engaging given the age, level, and background of her students. She must know how to group students for productive work by taking into account the dynamics of that particular collection of individuals in the class. She must be able to set up activities clearly and efficiently so that students know what they are doing and why, have enough language resources at their disposal to accomplish the task, and have sufficient time to complete the task without lollygagging. She must assure that students are attentive and on-task and make provisions for students who finish early or lag behind their classmates. These are all part of the craft of teaching that can determine whether a language class is a success that results in student learning or a failure that creates frustration for all concerned.

4 The Relationship between Theory and Practice

It is this kind of individual and particular experience of teachers that can alienate practicing teachers from the generalizations of theory as presented by university researchers. To overcome the "dysfunctions of the theory/practice discourse," Clarke (1994) recommends that teachers need to resist the imposition of experts' advice if it does not work for them. Conversely, Clarke advises theorists on university faculties to consider whether their work is not just generalizable but also "particularizable" to individual teachers. Crookes (1997) also questions the relationship between the research and theory of second language acquisition researchers and the concerns of second and foreign language teachers and suggests that the relationship could be improved if research focused on learning as social and was within a more qualitative tradition that would make it easier for practicing teachers to read.

Lightbown (2000) also asserts the need for dialogue between researchers and teachers. On the one hand, teachers should know what research has found, but on the other hand, researchers should listen to what teachers say. As a bridge between theory and practice, Lightbown (also echoed in Long 2011) lists a number of findings from second language acquisition research that she believes can have a direct bearing on the practice of classroom teaching. Group and pair work are useful techniques to enable students to get input and negotiate meaning. Although copious communicative practice is useful, and comprehensible input is necessary, neither practice nor comprehensible input is sufficient to produce both fluent and accurate use of the L2 by learners. Within a content-rich context, a focus on form is helpful in getting students to notice and use new and/or correct forms. Error correction is effective if it is sustained and focuses on something the learner is ready to acquire. Indeed, in terms of readiness for acquisition, learners progress through developmental stages, but their progress through a given stage can be speeded up by form-focused instruction.

In an effort to make research more accessible to practicing teachers, Macaro (2003) surveyed secondary school teachers of modern languages in England on research issues that they felt would be most useful for their practice. Of the 80 responses he received, top-ranking areas of interest seem to reflect day-to-day chalkface concerns. The first area was vocabulary acquisition

and what memorization techniques would prove most effective for learners. The second area of concern, how rules of grammar can best be learned, reflects the mandate of the 1999 National Curriculum in the UK, which called for students to be taught grammar and how to apply it. One of the most frequently mentioned areas of concern for teachers was the vexed issue of motivation, particularly for boys, slow or reluctant learners, and learners with xenophobic attitudes. Among the four skills (speaking, listening, reading, and writing), speaking garnered most interest. As Macaro notes, given 20 years of CLT in the UK and abundant interaction in modern language classes, it is of concern to teachers that the results are no better than they are.

5 English as Lingua Franca: A Challenge

Perhaps the poor results for language learning in the UK and the USA may be attributed at least in part to the rapid growth of English as lingua franca. Twenty-first-century globalization and the explosion of English language learning worldwide mean that we must look at language teaching and learning from a new perspective. At least one estimate counts a total of 2 billion speakers of English throughout the globe, with the number of non-native speakers of English outnumbering native speakers of English by three or four to one (Crystal 2008). Previously, it was relatively easy to categorize English language speakers according to whether they belonged to one of three "circles" proposed by Kachru (1986). The Inner Circle included those countries where English was the native language of the vast majority of speakers (e.g., the UK, the USA, Canada, Australia, New Zealand). In the Outer Circle, comprising mostly countries that had been former British colonies (e.g., Nigeria and India), English was a second language used as a link language within the country but not the native language of most speakers. The Expanding Circle included all those countries where English was a foreign language not used within the community (e.g., Turkey, China, countries in continental Europe). Now, however, these easy distinctions have blurred. Learners of English as an international language (EIL) or English as a lingua franca (known by the endearing acronym ELF) use the language to communicate with other non-native speakers of English in a variety of domains. With its wide geographical distribution and range of functions, English has become the lingua franca for science and technology, business, and higher education, to name just a few (McArthur 2003; Bolton 2008; McKay 2003).

English, then, is no longer "owned" solely by native speakers within the Inner Circle of Great Britain, the United States, Australia, and other countries that owe their origin to settlement by the British. As a result, ELF has become de-linked from Inner Circle native speaker norms of pronunciation, grammar, and pragmatics (rules for using the language within the social context). This calls into question our ideas of language connected to a native speaker community and the norms of language as a result. On the one hand, Seidlhofer (2001a: 146, 149) has described ELF as a language without native speakers, and thus a language for which there can be no native speaker intuitions to guide what is "correct." On the other hand, Canagarajah (2007: 925–6) points to the hybrid nature of the ELF speech community and asserts that *all* users of ELF have native speaker competence in it. Because the form of the language is fluid and negotiated anew with each encounter, it is a language with variable form and no real expectations of norms, depending as it does on the speakers in any given situation.

Not all those who have examined ELF would agree with Canagarajah's position. Instead, it has been asserted that in places like Europe and Asia, a localized variety of ELF has developed. In Europe, for example, the development of "Euro-English" has occasioned study and description not only of how the language is used but also how it can be described linguistically with an eye toward establishing a teachable standard. Jenkins (2002) has sought to describe the phonology of ELF that should be taught to learners. The core features were selected as core because they are necessary for comprehensibility between non-native speakers, such as: the rhotic *r* (pronunciation of the *r* after vowels in words such as *car* or *bir*th); aspiration of voiceless stops (producing a

little puff of air after the sounds [p], [t], [k] at the beginnings of syllables); preservation of consonant clusters word-initially; and the tense–lax distinction of vowels (e.g., the distinction between vowel sounds in words such as *r*i*ch* versus *r*ea*ch*, or *m*a*ce* versus *m*e*ss*). Noncore features, those not considered necessary for comprehension, include: the *th* sounds [θ,ð]; weak and reduced forms (such as swallowing the vowel sound in words such as *a*nd or *t*o in rapid speech); and stress-timed rhythm (evenly timing the stresses in a sentence regardless of the number of intervening syllables). In terms of grammar, Seidlhofer (2001b) has listed the following forms that are unacceptable in native speakers' English but are evolving as acceptable and comprehensible in Euro-English: deletion of the third person singular *–s* on present tense verbs; omission of definite and indefinite articles; loss of the *who/which* distinction; and *isn't it?* used as a universal tag question.

In China, the notion of "China English," a localized variety of English "with Chinese characteristics" (Kirkpatrick and Xu 2002: 269) has been proposed by some as a standard. The phonology of China English is characterized by a reduction in vowel contrasts and syllable-timed rhythm, inter alia. In vocabulary, it includes Chinese loan words such as *Putonghua* 'Mandarin Chinese' and *fengshui* and direct translations from Chinese such as *the four modernizations* and *to get rich is glorious* – expressions for concepts relevant to modern Chinese culture. In grammar, there is lack of the plural and third person singular *–s*, an invariant tag question, and a weakening of the count/mass noun distinction. In addition, Asian pragmatic norms are in operation, such as delaying a request until after facework has been done and reasons for the request are given (Bolton 2008: 8–9; Kirkpatrick and Xu 2002: 270–4). Although the idea of China English as a localized standard has been floated for some time, it has not garnered support from business and political leaders or indeed from learners themselves (Bolton 2008: 9–10; Hansen 2007: 274–8; Kirkpatrick and Xu 2002: 275–7; Li 2007: 12–14). Attitudes, however, may be changing at least so far as pronunciation is concerned; a more recent large-scale survey revealed greater tolerance for Chinese-accented English, while maintaining strong preference for native speaker grammar as the model (He and Zhang 2010).

In both the examples of Euro-English and China English, the proposal is that the language is to be judged not by its congruence with a native speaker model but by the effectiveness of its communication within the international context. Indeed, as Llurda (2004: 316) points out, these standards at least in Europe are already emerging naturally without language planning and will mean that the idea of communicative competence needs to be re-examined (Llurda 2004: 317). Before the turn of the millennium, the native speaker model had been questioned in general, and the interest in ELF has brought those questions to the fore. As Cook (1999) points out, the native speaker model is based on a flawed concept of an idealized normative monolingual speaker; however, the L2 learner is by definition not a monolingual. Furthermore, if the L2 speaker is compared with a native speaker, there is no hope of success since the L2 learner will never become a native speaker; instead, the learner should be viewed positively as a successful L2 *user*, not as a failed native speaker. The native speaker model is thus a deficit model because it faults the nonnative speaker for lacking what he can never have. This deficit model has been decried by other scholars who note its inappropriateness when dealing with emerging forms of EIL (e.g., Jenkins 2002: 85; Li 2007: 11). Within the context of EIL, users of English may not want or need native speaker competence and may even reject native speaker norms (McKay 2003: 5–6).

In removing the native speaker norm from consideration of language form, there remains the vexed question of the relationship between language and culture. English, for good or for ill, is associated not only with access to international business and research but also with colonial history and hegemonic politics in the modern age. According to Modiano (2001), British English – particularly the prestige form known as RP – is associated with class stratification and ideas of *noblesse oblige* and the "white man's burden," while American English is associated with insensitive political and economic hegemony on the world stage. Because language learning in some senses requires the learner to redefine his or her identity, Modiano warns of the danger that "the learner's mind is colonized" (p. 164). Without pointing to specifics, Modiano calls for teachers to

identify those teaching practices that would "disenfranchise the learner" so that ELF can become a tool for establishing "a democratic basis for cross-cultural communication" (p. 171).

However, as Kramsch (2006) points out, the world has changed as we become multilingual, global citizens who use L2s not just with native speakers in their culture but with other global citizens. As we negotiate meaning, we must understand not just the linguistic structures but the intentions, power, status, speaking rights, pride, honor, and face needs of our interlocutors. The position of the nonnative speaker is one within the "global market of symbolic exchanges" (Kramsch 2006: 250). Indeed, users of ELF may be appropriating the language variety to signify their own identity as global citizens. Similar to the way that global hip-hop is interpreted as performance of a global identity (Pennycook 2009), there is evidence from work on motivation that acquisition of ELF has important symbolic meaning that signifies the user's international posture (Kormos, Kiddle, and Csizér 2011). This new ELF reality will require new ways of thinking about pedagogy.

If the native speaker is not the model, native speaker competence is not the goal, and the language should be de-linked from inner circle cultural associations, what then is the role of the native speaker teacher? As Liu (1999: 97–101) points out, native speakers have been privileged as teachers of English and often given preference in hiring. However, by one estimate, 80 percent of teachers of English are non-native speakers of the language (McKay 2003: 8–9). It may be time for these teachers to receive the recognition and respect they deserve. Llurda (2004) lists a number of contributions that the non-native speaker teacher can offer. Being bilingual themselves, non-native teachers understand the challenges their students face in learning the language. They can also counter claims of English dominance by maintaining their own multicultural perspective. They can act as cultural and linguistic mediators by presenting the "multifaceted reality" of L2 use and help learners express their identity in English. Furthermore, the teacher's ability to code-switch will be a powerful illustration of cultural identity and linguistic competence (Llurda 2004: 318–20). And the non-native speaker is also a role model of what is achievable (Cook 1999: 199–200).

For the classroom teacher of ELF, publications have not set forth much specific detail on pedagogy. Cook (1999: 196–204) suggests setting appropriate goals that emphasize the learner as a potential successful user of the L2. To that end, materials and activities should present models of competent L2 use. Teaching should be based on a description of L2 users, for example by basing intermediate goals on corpora of actual L2 use. In course books and materials, the L2 user can be presented within his or her own culture interacting with native speakers or other non-native speakers of the target language, rather than being portrayed only as a visitor in an English-speaking culture. Teaching methods should also incorporate and allow for use of students' L1, which can even include updated translation exercises to involve the learner as an "intercultural speaker." McKay (2003: 18–19) stresses the need for a sufficiently complex understanding of the diversity of teaching and learning, of how EIL is used, and of cross-cultural encounters and how English is used in these encounters. She cautions that learners may not want or need the full range of registers or native speaker-like pronunciation or pragmatics. Canagarajah (2007) suggests that teachers need to develop students' ability to negotiate and focus on communicative strategies rather than on linguistic form. They should recognize that proficiency is measured by versatility in language use; it is not what the user knows, but what the user is able to do with the language. Jenkins (2012) probably offers the most realistic assessment of the pedagogical implications of ELF: She asserts that teaching applications must come from the teachers themselves, not from researchers removed from the language classroom.

6 What Does the Future Hold?

We should be entering a brave new world. Globalization has created greater need than ever to connect with people who speak other languages, and easy travel and new forms of technology

have made this ever more possible. In fact, while the primary source of language learning may still be the classroom, it is fast being supplemented by various sorts of media. A survey of 8,830 young people in six European countries, for example, revealed that, on average, the young people acquired 57.55 percent of their English through school, 25.96 percent through the media, and 17.19 percent through other means (DeBot 2007: 274–5).

Technology is opening up new possibilities of distance learning and self-access learning both in the context of formal classroom learning and for more individualized web surfers (see Blake 2007 for a summary and the 2009 focus issue of *Modern Language Journal*). Students in foreign language classes can be brought together for various forms of online communication with each other, and learners can easily access authentic texts through the internet, thus gaining a richer exposure to language used for authentic communicative purposes (e.g., Levy 2009; Thorne, Black, and Sykes 2009). Technology, however, is not the magic bullet, as attested by even a cursory glimpse at earlier promises of technology, namely language labs in the 1960s (e.g., Brooks 1964: 189–98) or electronic media in the 1970s (e.g., Arendt 1970).

Within Inner Circle English-speaking countries, interest in foreign language learning remains disappointing. Present enrollments in foreign languages at the university level in the United States are dishearteningly low. Although enrollments increased 13 percent from 2002 to 2006 and another 6.6 percent between 2006 and 2009, the rate of enrollment is still only about half what it was in the mid-1960s, and the vast majority of enrollments are in the first and second-year courses with only a small minority of students continuing to advanced courses where they could gain real fluency (MLA 2007; MLA 2010). A survey of elementary and secondary schools in the USA in the decade 1997–2008 found a decline in the number of schools offering foreign language instruction and in the actual number and percentage of pupils enrolled in foreign language (Pufahl and Rhodes 2011). The situation is likewise disappointing in the British Isles. An extensive survey of foreign language knowledge in Europe revealed that over 60 percent of the respondents in Ireland and the UK could not hold a conversation in any language but the L1 (TNS Opinion and Social 2006). In the UK, Mitchell (2000) cites the conclusion that secondary school pupils do not show creative ability to use the L2 they have studied, nor do they apply what they have previously learned to new language situations. Phillipson (1998) even suggests that policymakers in education do not really see the necessity for native speakers of English to learn other languages.

The picture in continental Europe is more optimistic; a survey of 28,694 Europeans in 2005 found that 56 percent could conduct a conversation in a language other than the L1, and 83 percent considered language learning useful for their lives (TNS Opinion and Social 2006). The role of English, however, is cause for concern. In Europe, there is a "growing tendency for 'foreign language learning' to mean simply 'learning English'," and the trend toward using English as the medium of instruction "may have unforeseen consequences" for national and regional languages (Commission of the European Communities 2005: 4, 6). In other parts of the world as well as Europe, there is concern that the spread of English is due to linguistic imperialism, that English is a "killer language" threatening indigenous languages and taking away the incentive to study other world languages (e.g., see Phillipson 1998; Phillipson and Skutnabb-Kangas 1997).

The success of L2 pedagogy can rest on factors such as the congruence of a teacher's theoretical understanding of language, second language acquisition, and practice within the classroom. Yet there are sociopolitical concerns that crucially affect students' attitudes and levels of motivation. While a skillful L2 teacher can do much to increase the positive attitudes of her students and their level of motivation to learn, the wider sociocultural context within which teaching takes place may have a far-reaching effect on the extent to which success is possible.

NOTE

1 The term *methodology* has been deliberately chosen to avoid the fine semantic distinctions in the academic literature among *approach*, *method*, and *technique* (e.g., Brown 2002; Richards and Rogers 2001) and to reflect teachers' less narrowly defined uses of the term *methods* (e.g., in Bell's 2007 study of teacher beliefs).

REFERENCES

Adamson, B. (2006). Fashions in language teaching methodology. In A. Davies and C. Elder (eds), *The Handbook of Applied Linguistics* (pp. 604–62). Malden, MA: Blackwell.

Allwright, D. (2005). Developing principles for practitioner research: the case of exploratory Practice. *Modern Language Journal*, 89, 3, 353–66.

American Council for the Teaching of Foreign Languages. (1983). *ACTFL Proficiency Guidelines.* Revised 1985. Hastings-on-Hudson, NY: ACTFL Materials Center. Also available online at www.sil.org/lingualinks/languagelearning/OtherResources/ACTFLProficiencyGuidelines/contents.htm

Arendt, J. D. (1970). Media in foreign language teaching. In D. L. Lange (ed.), *Britannica Review of Foreign Language Education*, vol. 2 (pp. 157–89). Chicago, IL: Encyclopedia Britannica and American Council on the Teaching of Foreign Languages.

Badger, R. and MacDonald, M. (2010). Making it real: authenticity, process, and pedagogy. *Applied Linguistics*, 31, 4, 578–82.

Barton, F. B. and Sirich, E. H. (1954). *Practical French Review: Grammar and Composition*. New York: Appleton-Century-Crofts.

Bell, D. M. (2003). Method and postmethod: are they really so incompatible? *TESOL Quarterly*, 37, 2, 325–36.

Bell, D. M. (2007). Do teachers think that methods are dead? *ELT Journal*, 61, 2, 135–43.

Blake, R. J. (2007). New trends in using technology in the language curriculum. *Annual Review of Applied Linguistics*, 27, 76–97.

Bolton, K. (2008). English in Asia, Asian Englishes, and the issue of proficiency. *English Today*, 94, 24, 2, 3–12.

Breen, M. P., Hird, B., Milton, M., Oliver, R., and Thwaite, A. (2001). Making sense of language teaching: teachers' principles and classroom practices. *Applied Linguistics*, 22, 4, 470–501.

Brooks, N. (1964). *Language and Language Learning: Theory and Practice*, 2nd edn. New York: Harcourt, Brace, and World.

Brown, H. D. (2002 [1997]). English language teaching in the "post-method" era: toward better diagnosis, treatment, and assessment. In J. Richards and W. A. Renandya (eds), *Methodology in Language Teaching: An Anthology of Current Practice* (pp. 9–18). Cambridge: Cambridge University Press.

Bruton, A. (2011). Are the differences between CLIL and non-CLIL groups in Andalusia due to CLIL? A reply to Lorenzo, Casal, and Moore (2010). *Applied Linguistics*, 32, 2, 236–41.

Butler, Y. G. (2011). The implementation of communicative and task-based language teaching in the Asia-Pacific region. *Annual Review of Applied Linguistics*, 31, 36–57.

Canagarajah, S. (2007). Lingua franca English, multilingual communities, and language acquisition. *Modern Language Journal*, 91, 923–39.

Carless, D. (2004). Issues in teachers' reinterpretation of a task-based innovation in primary schools. *TESOL Quarterly*, 38, 4, 639–62.

Clarke, M. (1994). The dysfunctions of the theory/practice discourse. *TESOL Quarterly*, 28, 1, 9–26.

Commission of the European Communities. (2005). A new framework strategy for multilingualism. Brussels: CEC. Available online at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0596:FIN:EN:PDF

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly*, 33, 2, 185–209.

Cortazzi, M. and Jin, L. (1996). Cultures of learning: language classrooms in China. In H. Coleman (ed.), *Society and the Language Classroom* (pp. 169–206). Cambridge: Cambridge University Press.

Crookes, G. (1997). Second language acquisition and language pedagogy: a socioeducational

perspective. *Studies in Second Language Acquisition*, 20, 93–116.
Crystal, D. (2008). Two thousand million? Updates on the statistics of English. *English Today*, 24, 1, 3–6.
Dalton-Puffer, C. (2011). Content-and-language integrated learning: from practice to principles? *Annual Review of Applied Linguistics*, 31, 182–204.
DeBot, K. (2007). Language teaching in a changing world. *Modern Language Journal*, 91, 274–6.
Dickens, P. M. R. and Woods, E. G. (1988). Some criteria for the development of communicative grammar tasks. *TESOL Quarterly*, 22, 4, 623–46.
Ellis, R. (2009). Task-based language teaching: sorting out the misunderstandings. *International Journal of Applied Linguistics*, 19, 3, 221–46.
European Commission. (2006). European language policy and CLIL: a selection of EU-funded projects. Brussels: Directorate-General for Education and Culture. Available online at http://ec.europa.eu/education/languages/pdf/doc_786_en.pdf
Garner, L. C., Rusch, D., and Domínguez, M. (2008). *Claro Que Sí!*, 6th edn. Boston, MA: Houghton Mifflin.
Gass, S. and Selinker, L. (2001). *Second Language Acquisition: An Introductory Course*, 2nd edn. Mahwah, NJ: Erlbaum.
Gower, R. and Walters, S. (1983). *Teaching Practice Handbook: A Reference Book for EFL Teachers in Training*. London: Heinemann.
Hansen Edwards, J. G. 2007. Teaching pronunciation in twenty-first century China: models and methods. In J. Liu (ed.), *English Teaching in China: New Approaches, Perspectives, and Standards* (pp. 270–84). London: Continuum.
He, D. and Zhang, Q. (2010). Native speaker norms and China English: from the perspective of learners and teachers in China. *TESOL Quarterly*, 44, 4, 769–89.
Hiep, P. H. (2007). Communicative language teaching: unity within diversity. *ELT Journal*, 61, 3, 193–201.
Hymes, D. (1979 [1972, 1966]). On communicative competence. In C. J. Brumfit and K. Johnson (eds), *The Communicative Approach to Language Teaching* (pp. 5–26). Oxford: Oxford University Press.
Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an International Language. *Applied Linguistics*, 23, 1, 83–103.
Jenkins, J. (2012). English as Lingua Franca from the classroom to the classroom. *ELT Journal*, 66, 4, 486–94.
Kachru, B. (1986). The power and politics of English. *World Englishes*, 5, 2/3, 13–19.
Kirkpatrick, A. and Xu, Z. (2002). Chinese pragmatic norms and "China English." *World Englishes*, 21, 2, 269–79.
Klee, C. A. (2009). Internationalization and foreign languages: the resurgence of interest in languages across the curriculum. *Modern Language Journal*, 93, 4, 618–21.
Kormos, J., Kiddle, T., and Csizér, K. (2011). Systems of goals, attitudes, and self-related beliefs in second language learning motivation. *Applied Linguistics*, 32, 5, 495–516.
Kramsch, C. (2006). From communicative competence to symbolic competence. *Modern Language Journal*, 90, 2, 249–52.
Krashen, S. D. and Terrell, T. D. (1988). *The Natural Approach: Language Acquisition in the Classroom*. New York: Prentice Hall.
Kumaradivelu, B. (1994). The postmethod condition: (e)merging strategies for second/foreign language teaching. *TESOL Quarterly*, 28, 1, 27–48.
Kumaradivelu, B. (2001). Toward a postmethod pedagogy. *TESOL Quarterly*, 35, 4, 537–60.
Lado, R. (1964). *Language Teaching: A Scientific Approach*. New York: McGraw-Hill.
Larsen-Freeman, D. (2011). Teaching and testing grammar. In M. H. Long and C. J. Doughty (eds), *The Handbook of Language Teaching* (pp. 518–42). Malden, MA: Wiley Blackwell.
Levy, M. (2009). Technologies in use for second language learning. *Modern Language Journal*, 93, S1, 769–82.
Li, D. C. S. (2007). Researching and teaching China and Hong Kong English. *English Today*, 23, 3/4, 11–17.
Li, D. (1998). "It's always more difficult than you plan and imagine": Teachers' perceived difficulties in introducing the communicative approach in South Korea. *TESOL Quarterly*, 32, 4, 677–703.
Lightbown, P. (2000). Anniversary article: classroom second language acquisition research and second language teaching. *Applied Linguistics*, 21, 4, 431–62.
Liu, J. (1999). Nonnative-English-speaking professionals in TESOL. *TESOL Quarterly*, 33, 1, 85–102.
Llurda, E. (2004). Non-native-speaker teachers and English as an International Language. *International Journal of Applied Linguistics*, 14, 3, 314–23.
Long, M. H. (2011). Methodological principles for language teaching. In M. H. Long and C. J. Doughty (eds), *The Handbook of Language Teaching* (pp. 373–94). Malden, MA: Wiley Blackwell.
Lorenzo, F., Casal, S., and Moore, P. (2010). The effects of content and language integrated learning in European education: key findings from the Andalusian bilingual sections evaluation project. *Applied Linguistics*, 31, 3, 418–42.
Lyster, R. and Ballinger, S. (2011). Content-based language teaching: convergent concerns across divergent contexts. *Language Teaching Research*, 15, 3, 279–88.

Macaro, E. (2003). *Teaching and Learning a Second Language*. London: Continuum.

Magnan, S. S. (2007). Reconsidering communicative language teaching for national goals. *Modern Language Journal*, 91, 249–52.

Matthews, A., Spratt, M., and Dangerfield, L. (eds) (1985). *At the Chalkface: Practical Techniques in Language Teaching*. London: Edward Arnold.

McArthur, T. (2003). World English, Euro-English, Nordic English? *English Today*, 19, 1: 54–8.

McDonough, K. and Chaikitmongkol, W. (2007). Teachers' and learners' reactions to a task-based EFL course in Thailand. *TESOL Quarterly*, 41, 1, 107–32.

McKay, S. L. (2003). Toward an appropriate EIL pedagogy: re-examining communicative ELT assumptions. *International Journal of Applied Linguistics*, 13, 1, 1–22.

Mitchell, C. B. and Vidal, K. E. (2001). Weighing the ways of the flow: twentieth century language instruction. *Modern Language Journal*, 85, 1, 26–38.

Mitchell, R. (2000). Anniversary article: Applied linguistics and evidence-based classroom practice: the case of foreign language grammar pedagogy. *Applied Linguistics*, 21, 3, 281–303.

MLA (Modern Language Association). (2007). New MLA survey shows significant increases in foreign language study at US colleges and universities. Press release available online at www.mla.org/pdf/release11207_ma_feb_update.pdf

MLA (Modern Language Association). (2010). Enrollments in languages other than English in United States institutions of higher education, Fall 2009. Available online at www.mla.org/pdf/2009_enrollment_survey.pdf

Modiano, M. (2001). Ideology and the ELT practitioner. *International Journal of Applied Linguistics*, 11, 2, 159–73.

Norris, J. M. (2011). Task-based teaching and testing. In M. H. Long and C. J. Doughty (eds), *The Handbook of Language Teaching* (pp. 578–94). Malden, MA: Wiley Blackwell.

Pennycook, A. (2009). Refashioning and performing identities in global hip-hop. In N. Coupland and A. Jaworski (eds), *The New Sociolinguistics Reader* (pp. 326–40). New York: Palgrave Macmillan.

Phillipson, R. (1998). Globalizing English: are linguistic human rights an alternative to linguistic imperialism? *Language Sciences*, 20, 1, 101–12.

Phillipson, R. and Skutnabb-Kangas, T. (1997). Linguistic human rights and English in Europe. *World Englishes*, 16, 1, 27–43.

Pica, T., Lincoln-Porter, F., Paninos, D., and Linnell, J. (1996). Language learners' interaction: how does it address the input, output, and feedback needs of language learners? *TESOL Quarterly*, 30, 1, 59–84.

Prabhu, N. (1987). *Second Language Pedagogy*. Oxford: Oxford University Press.

Pufahl, I. and Rhodes, N. (2011). Foreign language instruction in US schools: results of a national survey of elementary and secondary schools. *Foreign Language Annals*, 44, 2, 258–88.

Rice, F. A. and Sa'id, M. F. (1966 [1954]). *Eastern Arabic: An Introduction to the Spoken Arabic of Palestine, Syria, and Lebanon*. Beirut: Khayats.

Richards, J. C. and Lockhart, C. (1994). *Reflective Teaching in Second Language Classrooms*. Cambridge: Cambridge University Press.

Richards, J. C. and Renandya, W. A. (eds) (2002). *Methodology in Language Teaching: An Anthology of Current Practice*. Cambridge: Cambridge University Press.

Richards, J. C. and Rogers, T. S. (2001). *Approaches and Methods in Language Teaching*, 2nd edn. Cambridge: Cambridge University Press.

Savignon, S. J. (2001). Communicative language teaching for the twenty-first century. In M. Celce-Murcia (ed.), *Teaching English as a Second or Foreign Language*, 3rd edn (pp. 13–28). Boston, MA: Heinle and Heinle.

Schulz, R. A. (2006). Reevaluating communicative competence as a major goal in postsecondary language requirement courses. *Modern Language Journal*, 90, 252–5.

Seidlhofer, B. (2001a). Closing a conceptual gap: the case for a description of English as a lingua franca. *International Journal of Applied Linguistics*, 11, 2, 133–58.

Seidlhofer, B. (2001b). Towards making "Euro-English" a linguistic reality. *English Today*, 68, 14–16.

Shamim, F. (1996). Learner resistance to innovation in classroom methodology. In H. Coleman (ed.), *Society and the Language Classroom* (pp. 105–21). Cambridge: Cambridge University Press.

Sheen, R. (1994). A critical analysis of the advocacy of the task-based syllabus. *TESOL Quarterly*, 28, 1, 127–51.

Spada, N. and Lightbown, P. M. (2008). From-focused instruction: isolated or integrated? *TESOL Quarterly*, 42, 2, 181–207.

Spring, M. K. (2012). Languages for specific purposes curriculum in the context of Chinese-language flagship programs. *Modern Language Journal*, 96, S1, 140–57.

Stryker, S. B. and Leaver, B. L. (1997). Content-based instruction: from theory to practice. In S. B. Stryker and B. L. Leaver (eds), *Content-Based Instruction in Foreign Language Education: Models and Methods* (pp. 2–28). Washington, DC: Georgetown University Press.

Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible

output in its development. In S. Gass and C. Madden (eds), *Input in Second Language Acquisition* (pp. 235–53). Rowley, MA: Newbury House.

Swan, M. (2005). Legislation by hypothesis: the case of task-based instruction. *Applied Linguistics*, 26, 3, 376–401.

Tan, M. (2011). Mathematics and science teachers' beliefs and practices regarding the teaching of language in content learning. *Language Teaching Research*, 15, 3, 325–42.

Thorne, S., Black, R., and Sykes, J. (2009). Second language use, socialization, and learning in internet interest communities and online gaming. *Modern Language Journal*, 93, S1, 802–21.

TNS Opinion and Social. (2006). Europeans and their languages: summary. *Eurobarometer* 243/ *Wave* 64, 3. Brussels: European Commission. Available online at ec.europa.eu/public_opinion/archives/ebs/ebs_243_sum_en.pdf

Trim, J. L. M. (1979 [1973]). Draft outline of a European unit/credit system for modern language learning by adults. In C. J. Brumfit and K. Johnson (eds), *The Communicative Approach to Language Teaching* (pp. 100–2). Oxford: Oxford University Press.

Ur, P. (2013). Language-teaching method revisited. *ELT Journal*, 67, 4, 468–74.

Van den Branden, K. (2006). Introduction: task-based teaching in a nutshell. In K. Van den Branden (ed.), *Task-Based Language Education: From Theory to Practice* (pp. 1–16). Cambridge: Cambridge University Press.

van Ek, J. (1979 [1975]). The threshold level. In C. J. Brumfit and K. Johnson (eds), *The Communicative Approach to Language Teaching* (pp. 103–16). Oxford: Oxford University Press.

Waters, A. (2012). Trends and issues in ELT methods and methodology. *ELT Journal*, 66, 4, 440–9.

Wette, R. (2009). Making the instructional curriculum as an interactive, contextualized process: case studies of seven ESOL teachers. *Language Teaching Research*, 13, 4, 337–65.

Widdowson, H. G. (1979 [1972]). The teaching of English as communication. In C. J. Brumfit and K. Johnson (eds), *The Communicative Approach to Language Teaching* (pp. 117–21). Oxford: Oxford University Press.

Willis, J. R. (2004). Perspectives on task-based instruction: understanding our practices, acknowledging different practitioners. In B. L. Leaver and J. R. Willis (eds), *Task-Based Instruction in Foreign Language Education: Practices and Programs* (pp. 3–44). Washington DC: Georgetown University Press.

29 Educational Linguistics

JAMES PAUL GEE

1 Applied Linguistics and Theories of Language

Educational linguistics, as a branch of applied linguistics, varies based on the theory of language and language acquisition that one holds. One crucial area where theories of language differ is in how they see the relationship between language structure and language function. Another area where they differ is in how they view the relationship of areas like discourse and pragmatics to the more formal parts of linguistic theory such as the theory of syntax and logical form, and, indeed, whether they have a developed theory of discourse and pragmatics or not. And yet another crucial area where they differ is in the importance and role they assign to a biological basis for language acquisition.

Functional theories of language, such as Halliday's Systemic Functional Grammar (Halliday 1994; Martin 1992; Thompson 1996), which have informed a good deal of educational linguistics, especially outside the USA, see a very close and direct match or fit between language structure and language function, a match that has developed through history as language is used in different cultures and different social practices (e.g., science) for different purposes. Functional theories tend to view discourse and pragmatics as the fundamental level of language, since language use (the functions of language) shapes both the structure and meaning of language. Functional theories downplay the role of biology in language acquisition and tend to see all language acquisition, whether a child's first language or the child's acquisition of specialist varieties of language and literacy at school, as a process of socialization or enculturation into meaningful sociocultural practices. Acquiring a language or language practice is part and parcel of acquiring a culture or a sociocultural practice (Halliday 1993).

Generativist theories of language, such as Chomsky's various theoretical elaborations (e.g., Chomsky 1988, 1995), see no very close match between language structure and language function, structure being relatively autonomous from the functions it is used to serve. Generativist theories see syntax and logical form as the heart of language structure, amenable to scientific study, and see much of discourse and pragmatics as a branch of literary or social science and much less amenable to formal study. First language acquisition is heavily affected by a biological endowment for language and, thus, quite different from the later acquisition of specialist forms of language and literacy that rely much more heavily, indeed, in all likelihood, entirely, on learning and overt instruction (see Gee 1994, 2004 for discussion).

Let me take one example where these differing theories have played a major role in an important educational debate. An analogy between language acquisition (a type of immersion

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

learning) and other forms of later learning has been at the heart of many progressive pedagogies (see Cazden 1972: 139–42 for early and critical discussion). Progressive pedagogies, in their late twentieth-century versions, at least, stressed immersion in hands-on practice in meaningful environments where oral and written language are functional for the learners. Such pedagogies downplay the role of overt instruction. Such movements as "Whole Language" (and the earlier "Language Experience" approach) and "Process Writing" are examples of progressivist pedagogies (Goodman 1986, 1993, see Edelsky 1996 for discussion).

Progressivists tend to make the following sort of argument: Children acquire their native languages not by direct instruction (indeed, overt correction seems to have little impact on them), but by being immersed in rich, meaningful, and natural communicative settings. So, by analogy, it might be argued that in other areas, outside first language acquisition, humans learn best when their learning is self-motivated and self-directed in "natural" settings and not "imposed" on them by direct instruction.

This analogy between first language acquisition and later forms of language, literacy, and content acquisition has been attacked by a number of generative linguists who have argued that there are important aspects of first language acquisition that are not analogous to other forms of later learning (e.g., Melvold and Pesetsky 1995) . One of these involves biology. The child acquiring language is confronted with lots of "data" – the language she hears everywhere around her – though, of course, this data is always a very small subset of the infinite set of sentences in any language. There are always a great many hypotheses or patterns possible about what "rules" underlie the data (i.e., what generalizations there are in the data), especially granted the creativity of the human mind and the infinity of language. Therefore, something must "constrain" the child's "search space" such that the child does not "waste" exorbitant time considering fruitless or misleading hypotheses.

Generative linguists argue that the "something" that "helps" the child out here is in her genes. While there are debates as to what and how much of language is biologically specified, few doubt that the course of human evolution has made us humans "motivated" and "good at" acquiring our first languages (Bickerton 1981; Chomsky 1988; Elman 1991; Pinker 1994). For other sorts of learning – e.g., physics or literacy – evolution has not had enough time to build into human biology such a substantive and specific "step up," since things like writing and physics have simply not been around long enough in human history.

The point put (too) bluntly is this: Humans have an "instinct" for language and this greatly aids them in its acquisition (as certain species of birds have an instinct that helps them build their characteristic nests or sing their distinctive songs). We humans have no such instinct for acquiring any school subject like physics or literature, or for learning to read and write. As a result, there is far more variation in how people acquire these things than there is in how they acquire their first languages (the "failure rate," for instance, is dramatically different in the two cases).

The generative linguists' argument above about language and literacy addresses a very real problem. It, indeed, demands that progressivists develop a coherent theory of instructional guidance in the case of literacy and school subjects that can play (replace) the focusing ("scaffolding") role that human biology plays in first language acquisition (Gee 1994).

However, some people have taken the argument further than this and argued for a great deal of direct ("skill and drill") instruction in literacy education. This, of course, does not logically follow from the argument. All that follows is the need for some theory of teaching that can make up for the role that biology plays in the case of first language acquisition.

The argument for the biological basis of language is an argument about grammar, not everything else that goes under the rubric "language." To a generative linguist, "grammar" names the structural properties of sentences (phonological, morphological, syntactic, and logical). And it is about grammar in this sense that generative linguists make biological claims (and only grammar in this sense and, then, too, only the parts of grammar – parameters included – relevant to the basic design of all languages). Properties of meaning beyond semantics (as "type"

meaning, not meaning adjusted to contexts of use) and logical form and nearly all properties of "discourse" (both in terms of how language is put to use in context and in terms of how sentences are connected together to form "texts") do not fall under "grammatical theory" as generative linguists conceive it. But meaning and discourse are obviously crucial to later forms of language, content, and literacy learning and any recommendations we can make about teaching and learning in school (Gee 2011).

2 Social Languages

Children do not just learn a "native language" at home in their early socialization in life. They learn many different varieties or styles of language at school and throughout their lives. They may learn the "languages" of school-based talk and texts, mathematics, literary criticism, law, courts, street gangs, carpentry, and many more. Today, thanks to the growth of digital media and the ever greater sophistication of popular culture, children learn new varieties of language out of school as well as in, including quite technical varieties of language such as the language of *Yu-Gi-Oh*, video-game modding and theory-crafting, and language varieties that go with many forms of media production and citizen science. I will call such varieties of language – varieties fit for given tasks and social identities – "social languages" (Gee 2011, 2014 – the term "registers" is sometimes used in a similar way).

All social languages (whether the language of physicists or of street gangs) involve a mapping from the lexical, morphological, grammatical, and discourse resources of the language as a whole (e.g., "English" or "Spanish") to the distinctive grammatical and discourse features of the particular social language. This point sounds a bit arcane, so let me make it concrete with an example. Consider the following sentence (adapted from Halliday and Martin 1993):

Lung cancer death rates are clearly associated with an increase in smoking.

A whole bevy of linguistic features mark this sentence as part of a distinctive academic social language. Some of these are: a heavy subject ("lung cancer death rates"), deverbal nouns ("increase," "smoking"), a complex compound noun ("lung cancer death rates"), a low transitive relational predicate ("are associated with"), passive or passive-like voice ("are associated"), the absence of agency (no mention of who does the associating), an abstract noun ("rates"), and an assertive modifier to the verb ("clearly").

No one grammatical feature marks the social language of this sentence. Rather, all these features (and a great many more, including many discourse-level features in longer stretches of text) form a distinctive configuration – a correlation or, better, co-relation, that marks the social language. And, just like the mapping between phonemes and graphemes, this involves a form–form mapping (a mapping from formal features like deverbal nouns, heavy subject, passive voice, etc., to distinctive configuration of features that represents the social language).

This sort of form–form mapping is no easier for learners than the phonics (letter to sound mapping) one is. In fact, it is a good deal harder. Far more people fail to acquire a distinctive type of literacy because of a failure to acquire such "higher order" form–form mappings (mappings that lead to the recognition of social languages and genres of spoken and written language) than because of failing to acquire the phonics mapping.

Formal relationships like those we discussed in regard to the sentence discussed above do not exist, and are not learned, outside the distinctive social practices of which they are an integral part. They are part and parcel of the very "voice" or "identity" of people who speak and write and think and act and value and live that way (for a given time and place). To learn such relationships is part of what it means to learn to recognize the very social context one is in (and helping to create). Social languages can be acquired in a variety of different ways.

For the purposes of educational linguistics, it is best not to operate at the level of whole languages, but, rather, at the level of social languages. A given social language is composed of the

lexical and grammatical resources (whether in speech or writing) a recognizable group of people uses to carry out their characteristic social practices ("functions"). Los Angeles African American street gang members, laboratory physicists, medical doctors, successful middle school science students, and Los Angeles street policemen all have characteristic linguistic resources they use to carry out their distinctive social practices.

Distinctive social languages are constructed by social groups as tools to get special work done, whether this is playing *Yu-Gi-Oh* or doing physics. The "ways with words" of any social language are a matter of social conventions and conventions are not biologically specified but must be learned. While any social language uses the resources of a language like English or Spanish (whose core grammar may well be biologically specified), the conventions about how to use these resources for special tasks and practices are social and cultural.

Let me give an example to make matters clearer here. Biologists, and other scientists, write differently in professional journals than they do in popular science magazines. The popular science article is not merely a "translation" or "simplification" of the professional article. To see this consider the two extracts below, the first from a professional science journal, the second from a popular science magazine, both written by the same biologist on the same topic (example from Myers 1990: 150):

> 1. Experiments show that Heliconius butterflies are less likely to oviposit on host plants that possess eggs or egg-like structures. These egg-mimics are an unambiguous example of a plant trait evolved in response to a host-restricted group of insect herbivores. (professional journal)
>
> 2. Heliconius butterflies lay their eggs on Passiflora vines. In defense the vines seem to have evolved fake eggs that make it look to the butterflies as if eggs have already been laid on them. (popular science)

The first extract, from a professional scientific journal, is about the conceptual structure of a specific theory within the scientific discipline of biology. The subject of the initial sentence is "experiments," a methodological tool in natural science. The subject of the next sentence is "these egg mimics": note how plant parts are named, not in terms of the plant itself, but in terms of the role they play in a particular theory of natural selection and evolution, namely "coevolution" of predator and prey (that is, the theory that predator and prey evolve together by shaping each other). Note also, in this regard, the earlier "host plants" in the preceding sentence, rather than the "vines" of the popular passage.

In the second sentence, the butterflies are referred to as "a host-restricted group of insect herbivores," which points simultaneously to an aspect of scientific methodology (like "experiments" did) and to the logic of a theory (like "egg-mimics" did). Any scientist arguing for the theory of coevolution faces the difficulty of demonstrating a causal connection between a particular plant characteristic and a particular predator when most plants have so many different sorts of animals attacking them. A central methodological technique to overcome this problem is to study plant groups (like Passiflora vines) that are preyed on by only one or a few predators (in this case, Heliconius butterflies). "Host-restricted group of insect herbivores," then, refers both to the relationship between plant and insect that is at the heart of the theory of coevolution and to the methodological technique of picking plants and insects that are restricted to each other so as to "control" for other sorts of interactions.

The first passage, then, is concerned with scientific methodology and a particular theoretical perspective on evolution. On the other hand, the second extract, from a popular science magazine, is not about methodology and theory, but about animals in nature. The butterflies are the subject of the first sentence and the vine is the subject of the second. Further, the butterflies and the vine are labeled as such, not in terms of their role in a particular theory. The second passage is a story about the struggles of insects and plants that are transparently open to the trained gaze of the scientist. Further, the plant and insect become "intentional" actors in the drama: The plants

act in their own "defense" and things "look" a certain way to the insects, they are "deceived" by appearances as humans sometimes are.

These two examples replicate in the present what, in fact, is an historical difference. In the history of biology, the scientist's relationship with nature gradually changed from telling stories about direct observations of nature to carrying out complex experiments to test complex theories (Bazerman 1989). But here I want simply to note the ways in which the language of the professional passage generalizes over the sorts of events discussed in the language of the popular passage, and formalizes them within a specific set of (in this case, experimental) practices. To master the sorts of language used in the professional passage – whether orally or in writing – requires one not be able to understand and use language like that in the popular passage, but to have some meta-awareness of the relationship between these two forms of language – e.g., the different role of subject position in both.

And, we might add, the popular passage itself, as a form of "academic literacy" in its own right, represents a regimentation of yet more everyday forms of language – for example, the creation of technical terms that categorize things more "strictly" than we do in everyday language and the more specific and careful delineation of the grammar of argument form than is typical of everyday language (e.g., the chain "In defense" . . . "seem to have" . . . "that make it look" . . . "as if" . . .).

3 Equity

Educators have long worried about equity issues. There is ample evidence that children from poor and some minority homes do less well in school than do more advantaged children (Gee 2004). One reason such gaps exist is due to children's early socialization in life at home. Research has shown that the number of words a child has heard in conversation with an adult and the child's oral vocabulary at the age of five are two correlates with later success in school and society (Hart and Risley 1995). Extended talk with adults and hearing school- and book-based words from adults early in life appears to prepare children for the later complex language and cognitive demands of schooling. This is not meant to imply that children from poor families do not develop, at home early in life, complex linguistic and cognitive skills. It is meant only to say that some homes get their children ready for the specific sorts of styles of language – and their related ways of thinking – that will appear later in school and get ever more complex as schooling proceeds.

In regard to gaps between rich and poor in education an important issue arises over the nature of meaning in language. At the level of semantics, words have definition-like meanings. A word like "coffee" means a drink made from a particular type of plant (shrub) or just the plant itself. Such "literal" meanings actually set the possible range of meanings a word can have in context. In actual contexts of use, words have "situated meanings," contextually construed or contextually sensitive meanings. Thus, in "The coffee spilled, go get a mop," coffee means a liquid. In "The coffee spilled, go get a broom," it means a grain or bean. In "The coffee spilled, go stack it again," it means tins. And in "The coffee got smeared all over my shirt," it might mean coffee-flavored ice cream. We can understand "Big Coffee fights back" on analogy with "Big Oil fights back" and see "big coffee" as naming an industry.

On current theories of embodied cognition (Bergen 2012; see also Gee 1992) – closely related to older theories of localism or the locative hypothesis in linguistics and to new theories in cognitive linguistics (Lakoff and Johnson 1999) – humans think through the embodied experiences they have had. They use elements from these experiences to give situated meanings to words and language in use. On this theory, people only know, at a deep and useful level, what a specific social language means – such as the language of *Yu-Gi-Oh* or of physics – if they have had ample experiences in or with the world the language is about. Otherwise, their understandings are primarily verbal (definitional, literal). Of course, once people have a great deal of experience in

an area, they can learn more readily from texts and texts may be able, to some extent, to replace or actually be a form of experience in the domain.

These ideas from embodied cognition were in some ways anticipated by Vygotsky (1987). Vygotsky called concepts a person does not have reflective awareness of and conscious control over "spontaneous concepts." He called concepts over which a person does have reflective awareness of and conscious control over "nonspontaneous concepts."

Vygotsky argues that learners get nonspontaneous concepts only through working collaboratively with others who know more than they do and (simultaneously) via overt instruction that focuses on putting things into words, conscious and intentional use of the new concepts, and the relationships among forms and meanings. By working with others to carry out joint tasks that require the use of nonspontaneous concepts in such a way that the learner is focused consciously on conceptual connections, verbal links, and connections between form and meaning (not typical of everyday learning outside the classroom), the child comes to be able to use concepts in a reflective, controlled, and conscious way. This process eventually leads not just to the acquisition of some nonspontaneous concepts, but also to the transfer of this ability (conscious control and mastery) to the realm of everyday spontaneous concepts.

The child's everyday concepts become reorganized such that the child comes to realize the links and connections among her concepts; she comes to see them as, and to operate with them as, a system. The everyday spontaneous concept no longer hooks directly and singly to the world of experience, but hooks to experience now only via a whole network of relationships with other concepts.

To take one of Vygotsky's examples, the concept of "because" no longer hooks directly to concrete instances, but reaches reality now only through an intricate network of relationships to other concepts (e.g., physical causation, responsibility, contingency, various sorts of noncausal associations, and so forth). The child becomes aware that the concept has something of a life of its own and that it is related in various ways to other concepts. In turn, the child can now control her attributions of causal relationships with more reflection and mastery.

Vygotsky sums up his views as follows:

> The scientific concept is not related to its object directly. . . . this relationship is mediated by existing concepts.
>
> . . . the everyday concept acquires a whole series of new relationships with other concepts as it comes to stand between the scientific concept and its object. Its relationship with the object is also transformed in this process. (Vygotsky 1987: 223)

Vygotsky believes that it is because "scientific" school-based concepts were primarily invented and are learned at a conscious level and as part of an intricate system of links to other concepts that they can serve to develop the child's reflective abilities and reorganize the child's early conceptual development.

However, it is also clear that Vygotsky leaves us still having to delineate just what sorts of overt forms of instruction are fruitful. He clearly does not believe that rote drill or working on "irrelevant" materials is efficacious – he explicitly disowns such approaches (see Vygotsky 1987: 198–200). He does believe that the sorts of overt instruction that do work involve the learner actively working on problems with others so as to bring off accomplishments that she is not capable of alone. But, beyond that, Vygotsky clearly believes also that there are certain sorts of overt focusing on words, relationships, forms and functions, on conscious control and reflection, that must supplement mere collaborative problem solving – and we have not done well at cataloging these.

4 Language out of School

One of the most pressing issues in education today is the fact that many children from lower socioeconomic homes, many of them minority children, do poorly in school (Miller 1995). Such

children have often had little practice at home with school-based forms of language and interaction. Because of this, it is often assumed that they come to school with "nothing" relevant in the way of language practices on which to base their initiation into schooling.

However, in many cases, this is not true. Many of these children come to school well versed in language practices that, while not typical school-based literacy practices, are, nonetheless, rich verbal practices that can be leveraged to good effect by the school. Unfortunately, as I have said, the language practices of these children are often invisible to teachers and even, at times, denigrated. It should be a goal of schooling to allow such children to gain reflective awareness and conscious control over their own indigenous verbal practices as part of their acquisition of school-based social languages (in addition to their acquisition of "standard English").

School-based forms of language favor certain ways of seeing the world and certain ways of representing it. Schools, however, are often ignorant of or ignore styles of language that are home-based but less aligned with the school's ways with words. Rather than honoring and building on these practices, schools tend to denigrate them.

For example, sociolinguists have long known that the African American Vernacular English that some (but not all) African American children bring to school is a perfectly well-designed and logical dialect, no better or worse in grammatical terms than any other dialect (Baugh 1983; Labov 1972). Sociolinguists have also known for years that such children also bring to school richly intricate discourse practices, such as forms of storytelling (Smitherman 1977; Gee 2011). For instance, consider the story below, told by a seven-year-old African American girl ("Leona") in school at "sharing time" ("show and tell"). I have organized the story in terms of a line and stanza structure that helps to bring out its overall discourse organization:

Leona's Story
Frame
Stanza 1

1. Today
2. It's Friday the 13th
3. An' it's bad luck day
4. An' my grandmother's birthday is on bad luck day

Part 1: Making Cakes
Stanza 2

5. An' my mother's bakin' a cake
6. An' I went up my grandmother's house while my mother's bakin' a cake
7. An' my mother was bakin' a cheese cake
8. My grandmother was bakin' a whipped cream cupcakes

Stanza 3

9. An' we both went over my mother's house
10. An' then my grandmother had made a chocolate cake
11. An' then we went over my aunt's house
12. An' she had make a cake

Stanza 4

13. An' everybody had made a cake for nana
14. So we came out with six cakes

Part 2: Grandmother Eats Cakes
Stanza 5

15. Last night
16. My grandmother snuck out
17. An' she ate all the cake
18. An' we hadda make more

Stanza 6
(she knew we was makin' cakes)

19. An' we was sleepin'
20. An' she went in the room
21. An' gobbled em up
22. An' we hadda bake a whole bunch more

Stanza 7

23. She said mmmm
24. She had all chocolate on her face, cream, strawberries
25. She said mmmm
26. That was good

Stanza 8

27. An' then an' then all came out
28. An' my grandmother had ate all of it
29. She said "what's this cheese cake doin' here" – she didn't like cheese cakes
30. An' she told everybody that she didn't like cheese cakes

Stanza 9

31. An' we kept makin' cakes
32. An' she kept eatin' 'em
33. An' we finally got tired of makin' cakes
34. An' so we all ate 'em

Part 3: Grandmother Goes Outside the Home
Non-narrative section (35–41)
Stanza 10

35. An' now
36. Today's my grandmother's birthday
37. An' a lot of people's makin' a cake again
38. But my grandmother is goin' t'get her own cake at her bakery
39. An' she's gonna come out with a cake
40. That we didn't make
41. Cause she likes chocolate cream

Stanza 11

42. An' I went to the bakery with her
43. An' my grandmother ate cupcakes
44. An' an' she finally got sick on today
45. An' she was growling like a dog cause she ate so many cakes

Frame
Stanza 12

46. An' I finally told her that it was
47. It was Friday the thirteenth bad luck day

Leona's story is organized, through syntactic parallelism and lexical and syntactic repetition in an intricately poetic way. Note, for example, that every line in Stanza 2 ends on the word "cake" and that the stanza as a whole is organized by an aabb structure: ". . . bakin' a cake . . . bakin' a cake . . . bakin' a type of cake . . . bakin' a type of cake." Stanza 3 is organized in terms of an abab structure: "go over a house . . . make a cake . . . go over a house . . . bake a cake." Stanzas 5 and 6 are line by line stylistic variations of each other about the same event, an event which Stanza 7 mimes. Such poetic structuring, typical of African American oral storytelling, exists throughout the story.

The story also plays on a long-running theme in African American culture, namely the nature and ambiguities of symbols ("signifying") and the need for symbols to get the proper meaning in context (see Jackson 1974; Smitherman 1977; Stuckey 1987). The grandmother – the matriarch of the home – eats lots of cakes, big and small, at home and never gets sick. However, when she eats little cakes (cupcakes) outside the home at the bakery she does get sick and "growls like a dog" (loses her human status). What distinguishes humans from animals is their ability to see the symbolic value of things, to recognize for example, that a cake made at home is a true symbol of kinship, while one made at a bakery is not (it is simply a commodity), though it looks like one. The grandmother's momentary failure to recognize the nature of "real" and "duplicitous" symbols (signs) is what causes her to get sick and "growl like a dog."

Young African American children often help listeners interpret their stories by giving crucial non-narrative "evaluative" (Labov 1972) information right before the conclusion of the story (older storytellers spread such evaluative information throughout their stories). Leona does this is Stanza 10 where she stresses the importance of the fact that the grandmother is going to get a cake "that we [the family] didn't make."

Leona and other African American children in her class were often told to sit down during their sharing time turns, because the teacher felt they were not "talking about one important thing" (Michaels 1981), a rule of sharing time in the class. The teacher heard such children as meandering among different topics and places, loosely associating ideas and themes, since she, like many teachers, had little knowledge about the grammatical and discourse features of African American Vernacular English.

School often focuses children like Leona just on Standard English and not on their own dialect as well. However, if such children were to gain reflective and conscious control over a literary social language or a linguistic one as a "metalanguage" with which to reflect also on the grammatical and discourse properties of their own dialectal practices, they would gain deep knowledge about language, literature, and rhetoric (Lee 1993). And they would gain this knowledge in a way that links to their own cultural identity, thereby tying school, and school-based literacy practices, to that culture (rather than setting the two in resistance). After all, the sorts of poetic oral practices in which Leona is engaged, even as a seven-year-old, are historically at the

very root of literature in the West and elsewhere (e.g., the oral tradition from Homer to Chaucer). Through such practices, children like Leona could gain conscious control over the grammatical and discourse code of their community-based social language, as well as conscious control over school-based social languages connected to literature and literary criticism.

5 Linguistic Microanalysis

So far we have concentrated on language and not social interaction in classrooms. But many educational linguists focus on the ways in which language form and meaning, with concomitant consequences for empowering or disempowering people, are interactionally worked out moment by moment in specific educational contexts (Gee 2014; Green and Dixon 1993; Gumperz 1982; Lemke 1995). Discourse linguists have long realized that even what seem to be very minor aspects of language can have important consequences in communication. I will give one example here of how "minor" details can take on major importance in interaction, and, thus, must, in those instances, be included in linguistics transcripts (see Gee 2014). Consider the following interaction between a white female researcher ("R") and a fourth grade African American girl ("S" for student) with whom the researcher is discussing light as part of a school science education project:

1. R: Where does the light come from / when it's outside? //
2. S: Sun (low pitch) //
3. R: From the sun (low pitch) // .. hum
4. S: Cause the sun comes up / REALLY early //
5. R: um .. And that's when we get light (low pitch) //
6. S: And that's how the, the the me .. my .. me and my class / is talkin' about dinosau:rs / and how they die:d // And we found out .. / some things . about how they die:d //
7. R: Oh really // Does that have to do with LIGHT? //

In this transcript, a double slash ("//") indicates a final contour, that is, a rising or falling pitch of the voice that sounds final, as if a piece of information is closed off and finished (the fall or rise is realized over the underlined words and any words that follow them). A single slash ("/") indicates a nonfinal contour, a shorter rising or falling contour that sounds as if there is more information to come. Material between single or double slashes is called a tone unit (each tone unit has one word or phrase in it with major stress). Words that are capitalized carry the major stress in their tone unit (stress in English is marked by gliding the pitch of the voice up or down or increasing loudness or both). Capitalized words are emphatic (said with extra stress). Two periods (..) indicate a hearable pause. Two dots following a vowel ("die:d") indicate that the vowel is elongated (drawn out). "Low pitch" means that the preceding unit was said on overall low pitch. This transcript is certainly nowhere as narrow as it could be, though it includes some degree of linguistic detail.

After a long interaction from which this bit is taken, the researcher felt that the child often went off topic and was difficult to understand. However, it can be argued, from the above data, that the researcher co-constructed (contributed to) these topic changes and lack of understanding.

Children in school are used to a distinctive school activity in which an adult asks them a question (to which the adult already knows the answer, but to which the answer is not supposed to be obvious), the child answers, and the adult responds in some way that can be taken as evaluating whether the child's answer was acceptable or not (Sinclair and Coulthard 1975). In the above interaction, the researcher starts with a question to which the student responds with the word "sun" said on a low pitch and with a final falling contour. This way of answering indicates

(in many dialects of English) that the respondent takes the answer to be obvious (this already constitutes a problem with the question–answer–evaluation activity).

The researcher's response is said in exactly the same way as the child's (low pitch, final falling contour) – and in just the position that a student is liable to expect an evaluation – indicating that she, too, takes the answer to be obvious. The student might well be mystified, then, as to why the question was asked.

In (4) the student adds a tone unit that has an emphatic "really" in it and which is said on a higher pitch (basically on her normal level) and with a falling contour. This way of saying her contribution indicates that the student takes this information to be new or significant information. She may well have added this information in a search for some response that would render the initial question something other than a request for obvious or trivial information and in a search for some more energetic response from the researcher, one that would let the student know she was on the right tract in the interaction.

However, the student once again gets a response from the researcher (low pitch, falling final contour) that indicates the researcher takes the student's contribution, again, to be obvious. The student, then, in (6), launches off on yet another contribution that is, once again, said in a way that indicates she is trying to state new or significant information that will draw a response of interest from the researcher. The student also here uses a technique that is common to some African American students, though not to Anglo-American ones (Gee 1985): She states background information first before stating her main topic (light), though her "found out/some things" clearly implies, in this context, that these things will have to do with light (which they, indeed, do – she has studied how a meteor blocked out sunlight and helped destroy the dinosaurs). The researcher, listening for a more foregrounded connection to light, stops the student and, with emphasis on "light," clearly indicates that she is skeptical that the student's contribution is going to be about light, a skepticism that is, from the student's perspective, not only unmerited, but actually surprising and a bit insulting (as subsequent interaction shows).

Here the "devil" is, indeed, in the details: aspects of the school-based "known question-answer–evaluation" activity, different assumptions about how information is introduced and connected, as well as details of pitch and emphasis (as well as a good many other such details) all work together to lead to misunderstanding. This misunderstanding is quite consequential when the adult authority figure attributes the misunderstanding, not to the details of language and diversity (most certainly including her own language and diversity), but to the student herself.

One may wonder why the researcher asked the questions she did and responded as she did. To make a long story short, the research project was based on the idea that giving children too much explicit information or overt challenging responses would restrict their creativity and "sense making." Ironically, a situation set up to elicit the best from the child by leaving her as "free" as possible led to her being constructed as not making sense, when, in fact, she was making sense at several levels in a deeply paradoxical setting created by the researchers.

6 Conclusion

This chapter has ranged through a number of areas where knowledge about language and linguistics is relevant to current debates in education. Its basic points have been as follows:

1. Children do not just pick up school-based social languages and literacy through the sorts of rich immersion in socialization that is characteristic of first language acquisition. Teachers need to supplement such immersion, which is necessary but not sufficient for learning in school, with more overt forms of focusing on the structure of language and its complex relationships to communicative functions within different styles of language and texts.

2. An overt focus on social languages and specific genres of spoken and written language leads to conscious control of and meta-awareness about language that is fundamental to real understanding and that reorganizes how students think about their earlier forms of language and ways of thinking about the world. This is, indeed, one of the ultimate goals of schooling, and is particularly important for those children who come to school from homes that have not immersed them in school-based forms of language and interaction.
3. Schools often ignore or miss the resources of children who come from "nonmainstream" homes – homes which may not have immersed their children in school-based forms of language and interaction, but which have, nonetheless, immersed them in complex and culturally distinctive linguistic and interactional practices. Schools can honor the resources these children bring with them to school and build on them. In fact, they can allow such children to focus on form and function in their own styles of language in juxtaposition to other styles, especially those used in school, as a way to appreciate variation in language, respect their own culturally distinctive forms of language, and gain meta-awareness of how form and function work across different social languages in and out of school.
4. And, finally, teachers often assume that when a child appears to make little sense, especially a child from a different social and cultural group than their own, that the problem resides inside the child as a "deficit" of some sort. However, such problems often reside in the very interactions in which the teacher is taking part and in the teacher's lack of knowledge about the culturally distinctive resources different children bring with them to school. This dilemma calls both for better training for teachers in regard to language and linguistics (training which, in the USA, is now virtually nonexistent) and for teachers to engage in research on their own students and classrooms as a way to better understand the children they teach.

REFERENCES

Baugh, J. (1983). *Black Street Speech: Its History, Structure and Survival*. Austin, TX: University of Texas Press.

Bazerman, C. (1989). *Shaping Written Knowledge*. Madison, WI: University of Wisconsin Press.

Bergen, B. K. (2012). *Louder than Words: The New Science of How the Mind Makes Meaning*. New York: Basic Books.

Bickerton, D. (1981). *Roots of Language*. Ann Arbor, MI: Karoma.

Cazden, C. B. (1972). *Child Language and Education*. New York: Holt, Rinehart and Winston.

Chomsky, N. (1988). *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Edelsky, C. (1996). *With Literacy and Justice for All*, 2nd edn. London: Taylor and Francis.

Elman, J. (1991). Incremental learning, or the importance of starting small. *Technical Report 9101*. Center for Research in Language, University of California at San Diego.

Gee, J. P. (1985). The narrativization of experience in the oral style. *Journal of Education*, 167, 9–35.

Gee, J. P. (1992). *The Social Mind: Language, Ideology, and Social Practice*. New York: Bergin and Garvey.

Gee, J. P. (1994). First language acquisition as a guide for theories of learning and pedagogy. *Linguistics and Education*, 6, 331–54.

Gee, J. P. (2004). *Situated Language and Learning: A Critique of Traditional Schools*. London: Routledge.

Gee, J. P. (2011). *Social Linguistics and Literacies: Ideology in Discourses*, 4th edn. London: Taylor & Francis.

Gee, J. P. (2014). *An Introduction to Discourse Analysis: Theory and Method*, 4th edn. London: Routledge.

Green, J. and Dixon, C. (1993). Talking knowledge into being: discursive practices in classrooms. *Linguistics and Education*, 5, 231–9.

Goodman, K. (1986). *What's Whole in Whole Language?* Portsmouth, NH: Heinemann.

Goodman, K. (1993). *Phonics Phacts*. Portsmouth, NH: Heinemann.

Gumperz, J. J. (1982). *Discourse Strategies*. Cambridge: Cambridge University Press.

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education*, 5, 93–116.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*, 2nd edn. London: Edward Arnold.

Halliday, M. A. K. and Martin, J. R. (1993). *Writing Science: Literacy and Discursive Power*. Pittsburgh, PA: University of Pittsburgh Press.

Hart, T. and Risely, B. (1995). *Meaningful Differences in the Early Experience of Young American Children*. Baltimore, MD: Brookes.

Jackson, B. (1974). *"Get Your Ass in the Water and Swim Like Me": Narrative Poetry from Black Oral Tradition*. Cambridge, MA: Harvard University Press.

Labov, W. (1972). *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia, PA: University of Pennsylvania Press.

Lakoff G. and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Lee, C. D. (1993). *Signifying as a Scaffold for Literary Interpretation: The Pedagogical Implications of an African American Discourse Genre*. Urbana, IL: National Council of Teachers of English.

Lemke, J. (1995). *Textual Politics: Discourse and Social Dynamics*. London: Taylor & Francis.

Martin, J. R. (1992). *English Text: System and Structure*. Philadelphia, PA: John Benjamins.

Melvold, J. and Pesetsky, D. (1995). Reading is not like speaking. *The Boston Globe*, October 29 (Learning Section).

Michaels, S. (1981). "Sharing time:" Children's narrative styles and differential access to literacy. *Language in Society*, 10, 423–42.

Miller, L. S. (1995). *An American Imperative: Accelerating Minority Educational Advancement*. New Haven, CT: Yale University Press.

Myers, G. (1990). *Writing Biology: Texts in the Social Construction of Scientific Knowledge*. Madison, WI: University of Wisconsin Press.

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: HarperCollins.

Sinclair, J. M. and Coulthard, M. (1975). *Towards an Analysis of Discourse: The English Used by Pupils and Teachers*. Oxford: Oxford University Press.

Smitherman, G. (1977). *Talkin and Testifin: The Language of Black America*. Boston, MA: Houghton Mifflin.

Stuckey, S. (1987). *Slave Culture: Nationalist Theory and the Foundations of Black America*. New York: Oxford University Press.

Thompson, G. (1996). *Introducing Functional Grammar*. London: Edward Arnold.

Vygotsky, L. S. (1987). *The Collected Works of L. S. Vygotsky, Vol. 1. Problems of General Psychology. Including the Volume Thinking and Speech*, edited by R. W. Rieber and A. S. Carton. New York: Plenum.

30 Linguistics and Reading*

REBECCA TREIMAN

Much linguistic and psycholinguistic research has examined the structure and processing of spoken language. In this chapter, the focus changes to written language. The goal of the chapter is to review what is known about the processes involved in reading and in learning to read. How can skilled readers identify so many words so quickly? What is the best way to teach children to read and spell, and why do some otherwise normal children have such trouble learning to do so? Does knowledge of written language change people's basic intellectual or linguistic abilities? These and other questions will be addressed in the chapter.

Investigators from a variety of disciplines, including cognitive psychology, developmental psychology, and education, have contributed to research on reading. This mix reflects the fact that the study of reading is both theoretically interesting and practically important. Reading is a domain in which experimental psychologists can study fundamental questions such as how the knowledge stored in people's long-term memories affects their perception of the world. Reading is also a domain in which research findings have implications for important practical issues, such as classroom teaching and the diagnosis and treatment of learning problems. It is no wonder, then, that a large amount of research has been carried out on reading. Much of this research has been done with readers of English, but other languages have been examined too. The discussion begins with a consideration of the cognitive processes that are involved in skilled reading.

1 Bottom-Up and Top-Down Processing in Reading

Psychologists have distinguished two kinds of processing that are involved in reading and many other cognitive tasks. *Bottom-up processes* take in information from the outside world – symbols of a writing system, in the case of reading – and deal with that information with little recourse to higher-level knowledge. With *top-down* processes, on the other hand, uptake of information is guided by people's prior knowledge and expectations. In most situations, bottom-up and top-down processes work together to ensure rapid and accurate performance.

* Preparation of this chapter was supported in part by NIH grant HD051610 and NSF grant BCS-1421279.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

Theories about the cognitive processes involved in reading differ in the emphasis that they place on bottom-up and top-down processing. Theories that stress bottom-up processing focus on how readers extract information from the printed page, claiming that readers deal with letters and words in a relatively complete and systematic fashion (e.g., Gough 1972). Theories that stress top-down processing claim that readers form hypotheses about which words they will encounter and take in only just enough visual information to test their hypotheses. In the words of Goodman (1967), reading is a psycholinguistic guessing game. Although this view was originally propounded some time ago, its influence continues to be felt in some quarters today.

An example may help to clarify the distinction between theories that stress bottom-up processing and those that stress top-down processing. Suppose that someone has just read, *Daylight savings time ends tomorrow, and so people should remember to change their…* According to the top-down view, the reader guesses that the next word in the sentence will be *clocks*. The reader checks that the word begins with *c* and, because her hypothesis has been supported, barely takes in the remaining letters of the word. Theories of reading that stress bottom-up processing claim that the reader processes all of the letters in the last word of the sentence, regardless of the word's predictability.

Studies of how people's eyes move when they are reading texts provide some insight into the roles of bottom-up and top-down processes in reading. Research has shown that the eye does not sweep across each line of text in a continuous fashion. Rather, the eye comes to rest for somewhere around a quarter of a second, in what is called a *fixation*, and then makes a rapid jump (called a *saccade*) to the next fixation. It is during the fixation that visual stimulation is taken in; little or no useful information is extracted during a saccade. Research shows that skilled readers fixate at least once on the majority of words in a text. They do not skip a large number of words, as the top-down view predicts, but instead process the visual information rather thoroughly. Readers do this, in part, because their span of useful vision is fairly small. For example, a reader who fixates the *a* of *daylight* will probably be able to see all of the letters in this word. He may or may not be able to see enough to identify the next word, *savings*, but he will be unable to see the letters in *time*. Thus, the eye movement data portray reading as more of a bottom-up process than a top-down process. (Rayner et al. 2012 include a review of eye movement research.)

Comparisons of good and poor readers further support the claim that bottom-up processes play an important role in reading. If reading were a linguistically guided guessing game, as top-down theorists maintain, one would expect guessing ability to discriminate between good and poor readers. In that view, good readers are highly sensitive to context and use it to guide their uptake of print, whereas poor readers have trouble predicting the upcoming words in a sentence. However, research has shown that less skilled readers use context at least as much as skilled readers (Stanovich 1980). Skilled readers' perceptual skills are so accurate and automatic that they do not usually need to guess.

The statement that bottom-up processes play an important role in reading does not mean that top-down processes are unimportant. Words that are predictable from context are fixated for less time and are skipped more often than words that are less predictable, although the effects are relatively modest (Rayner et al. 2012). These results may be interpreted to mean that readers sometimes use their higher-order thinking skills to predict the upcoming words in a sentence. However, the findings may alternatively reflect low-level associative processes within the reader's *lexicon* or mental dictionary. For example, readers may spend less time on c*ake* in the sentence *The guests ate the wedding cake* than in the sentence *The guests ate the large cake* because *cake* is automatically activated once *wedding* has been recognized. Whatever mechanism is responsible for context effects, it is important to keep in mind that the words that contribute most to the meaning of a text are usually not very predictable from context. For example, almost any adjective or noun could follow *the* in a sentence. Readers must be able to use bottom-up processing if they wish to gain new information from what they read – the goal of reading – as opposed to reinforce what they already know.

2 Word Recognition

Many of the processes that are involved in understanding what we read are similar to the processes involved in understanding what we hear. With both written and spoken language, general knowledge about the world is often needed in order to make sense of and elaborate on the information that is presented. When reading about a wedding, for example, it is helpful to know about the kinds of activities that usually take place on such occasions. The grammatical knowledge that is necessary to understand a sentence is also similar whether the words are read or heard. Differences do arise in some cases. In English, for example, appositional phrases (as in the sentence *The wedding cake, which was made by a well-known baker, was beautiful*) occur more often in written language than in spoken language. In Arabic, the gap between the structures of written and spoken language is quite a bit larger than it is in English. What mainly distinguishes reading from speech, though, is the need to identify words by eye. Readers must recognize printed words accurately and automatically, linking the printed forms to the representations that are stored in the mental lexicon. This process of written word recognition has been a central focus of reading research.

To understand the processes that are involved in the recognition of printed words, it is important to consider how printed words map onto speech. Although writing systems differ from one another in many ways, all full writing systems are based on language (DeFrancis 1989; see Chapter 5). In a *logographic* system, each word (roughly speaking) has its own symbol. In other systems, the written forms of words reflect their sounds in a systematic manner. In some cases, these links are at the level of the syllable. For example, each syllable (roughly speaking) in spoken Japanese has its own symbol in the Japanese scripts called *kana*. In alphabetic writing systems, the link between print and speech is at the level of individual sounds or *phonemes*. Some alphabetic writing systems, such as Italian and Finnish, exemplify the alphabetic principle almost perfectly, with each letter representing one and only one phoneme. Other writing systems, including English and French, do not have such simple one-to-one links between phonemes and letters. For example, some English sounds have more than one possible spelling, as when the "k" sound is alternatively spelled as *c* (*cat*), *k* (*kit*), *q* (*quit*), or *ck* (*pack*). Also, some letters have more than one possible pronunciation. For example, *c* corresponds to "k" in the English *cabin* or the French equivalent *cabine*; it can also be "s" as in the English *circus* and the French equivalent *cirque*. Although such complications make the English and French writing systems more complex than some other alphabetic systems, they do not negate the value of the alphabetic principle. A skilled reader of English might pronounce *mook* to rhyme with *book* or *spook*, but he would never pronounce it as "vab." Certain deviations from the alphabetic principle are themselves principled, reflecting a tendency to spell *morphemes* (units of meaning) in a consistent fashion. For example, the English past tense ending is variously pronounced as "t" (as in *jumped*), "d" (as in *hemmed*), or "ud" (as in *wanted*), but in all three cases it is normally spelled as *ed*. As another example, the *a* in *health*, which makes the word exceptional from an alphabetic standpoint, reveals the relationship to *heal*. Other deviations from the alphabetic principle reflect a tendency to maintain the spellings of words that are borrowed from other languages. For example, French uses the un-Frenchlike spelling *sandwich*, which was borrowed from English.

Just as the printed forms of words reflect their linguistic forms, so the processing of printed words involves the recovery of the words' linguistic forms. Readers often access the phonological (or sound) forms of words as part of the recognition process. This phonological activation is covert, for skilled readers who are reading silently, but psychologists have devised clever ways to detect it. In one technique, people are presented with a category such as *type of food* and must then rapidly decide whether various printed words belong to the category. University students sometimes misclassify words that sound like category members (e.g., *meet*) as members of the category, even when they know the words' correct spellings. People make fewer errors on words

that look equally like a member of the category but that do not sound like one (e.g., *melt*) (Van Orden 1987). The results of these and other studies show that rapid, automatic word recognition is critical to reading success and that such recognition often involves activation of words' spoken forms. (See Frost 1998 for a review.)

There is some debate about exactly how readers derive the phonological forms of words from their spellings. Do skilled readers use explicit rules of the kind taught in phonics lessons, such as that *b* corresponds to "b," *sh* to "sh," and so on? Or do they rely on a network of implicit connections? Do readers use probabilistic patterns, such as that *ea* has a certain probability of corresponding to the "long e" sound (as in *steam*) and a lower probability of corresponding to the "long a" sound (as in *steak*), or do they use all-or-none rules that are not sensitive to a letter's context? Do the links between spellings and sounds operate only at the level of individual *graphemes* (letters and letter groups that correspond to single phonemes, such as *b* and *sh*) and phonemes? Or do people use rules that link sequences of letters, such as *ook*, to sequences of sounds?

Different theories about the spelling-to-sound translation process offer different answers to the questions laid out above. Theories of the *dual-route* type claim that skilled readers read words via both *lexical* and *nonlexical* routes. When using the lexical route, a reader looks up a word in her mental lexicon and accesses the pronunciation, if it has previously been stored there. When using the nonlexical route, the reader assembles a pronunciation using rules that, according to most dual-route models, relate individual graphemes to individual phonemes in an all-or-none manner. Dual-route theories claim that both the lexical and nonlexical routes are involved in the reading of many words. For example, a reader might retrieve the full pronunciation of *bun* from her mental lexicon while simultaneously gaining information about the word's pronunciation by combining the phonemes corresponding to *b*, *u*, and *n*. If the reader comes across an item that she has not previously encountered, such as *zun*, she uses the rules of the nonlexical route in order to decode it. Computer simulations of skilled readers that embody dual-route hypotheses have been developed for English and several other languages (Coltheart 2005; Ziegler, Perry, and Coltheart 2000).

Other models claim that a single route suffices for word recognition. Many of these models are *connectionist* in nature, seeking to explain single-word reading (and other aspects of cognition) in terms of networks of simple units that are connected to one another. Computer models of single-word reading include units that represent the input – the letters in a printed word and their order – and units that represent the output – the phonological form of the word. There are also *hidden units* that mediate between the input and output units. The computer model is taught to read by exposing it to pairs of printed and spoken words in a way that is thought to capture important aspects of a child's experience, including the fact that more common words are seen more often. The computer model generates a pronunciation for each word that is presented, compares it to the correct pronunciation, which is provided to it, and adjusts the weights of various connections so as to bring the generated pronunciation closer to the correct one. Over the course of many exposures to words, the weights on the model's connections begin to approximate the structure of the vocabulary on which it was trained. For example, if a model is taught the pronunciations of *bit*, *boot*, *book*, *boost*, and *brook*, the learned weights come to capture the fact that words beginning with *b* have pronunciations beginning with "b" and that words with medial *oo* have pronunciations that contain either the vowel of *boot* and *boost* or the vowel of *book* and *brook*, with the latter occurring before final *k*. The models are thus not restricted to simple, context-free links between graphemes and phonemes. They may assign one pronunciation to a letter or letter group when it occurs in one position of a word or when it is preceded or followed by a particular letter, and a different pronunciation when it occurs in another context. Models that follow these general principles have been developed for several languages (Ans, Carbonnel, and Valdois 1998; Seidenberg 2007).

Some studies (see Treiman and Kessler 2007 for discussion) suggest that phonological activation does not take place at only two levels – that of the whole word (the lexical route of the

dual-route model) and that of single graphemes (the nonlexical route). Patterns also exist at intermediate levels, not only in English but in other languages, and readers take advantage of these patterns. Readers are not limited to all-or-none rules that relate individual graphemes to individual phonemes and that do not take context into account. Given this, connectionist models have more promise than dual-route models as a way of explaining the performance of skilled readers. They may also help to explain the acquisition of word reading skills, the topic to which we now turn.

3 Learning to Read

In a number of English-speaking countries, including the United States, two general approaches have been put forward about how children should be taught to read (Rayner et al. 2002). The first of these, the *whole language* approach, is based on the idea that top-down processing plays an important role in reading. If skilled readers use context to predict the upcoming words in a sentence, only processing the print to the degree that is necessary to confirm their expectations, then children should do the same. Children should focus on the meaning of what they read rather than laboriously sounding out the individual words. Just as children will master spoken language if they are spoken to by others and given the opportunity to respond, so children will become literate if their environment is filled with print and if they are encouraged to explore it. Teachers who use a whole language approach thus focus on the meaning and purpose of printed language rather than on the individual letters and sounds. Activities may include reading stories to children and helping children use the pictures or the context to figure out the words. Sounding out an unknown word is typically considered a strategy of last resort, and children receive little guidance on how to do this. Whole language teachers encourage the integration of reading and writing, expecting children to write independently from an early age and offering little or no systematic instruction in conventional spelling.

A second class of approaches to literacy instruction, known as *phonics*, stresses the bottom-up processing of letters and words. In this view, learning to read and write is quite different from learning to understand and talk. Spoken language is deeply rooted in biological evolution and is as old as the human species itself. All normal members of the species learn to speak and understand without explicit tuition, provided that they are exposed to a spoken language. The situation is quite different for written language. Writing is a cultural achievement dating back some five or six thousand years; it is found among some groups of people but not others. Learning to read, advocates of phonics claim, usually requires explicit instruction. Children must learn to convert unfamiliar printed words into their familiar spoken forms by learning that *b* is pronounced as "b," that *c* may be pronounced as "k" or "s" depending on the identity of the following letter, and so on. This sounding out process is slow and laborious at first, but it becomes fast and automatic with practice. The phonics approach thus focuses on individual letters and sounds, repetition, and practice. Content and interest are not the only criteria for choosing reading materials; the words must also be easy to decode. For example, a story about a *bug* that eats *fish* would be preferred to a story about a *worm* that can *talk*, as *o* does not have its typical pronunciation in *worm* and *a* and *l* do not have their typical pronunciations in *talk*. Advocates of phonics maintain that the focus of initial reading instruction should be on the reading of individual words. Books designed for young children have simple grammar and vocabulary, and children will be able to understand the meaning if they can decode the words.

In practice, many programs include a blend of whole language and phonics activities. For example, children who are receiving phonics instruction learn about the meaning and function of print by reading (or being read) interesting stories. Use of writing for communication of meaning can be emphasized in phonics classrooms as well as in whole language classrooms. The central question is whether early reading instruction should include instruction in phonics. The answer

to this question, most researchers now agree, is that it should. Bottom-up skills are important for reading, as discussed earlier in the chapter, and most children need systematic instruction in order to develop these skills. Across a large number of studies, programs that include attention to phonics tend to yield better results than programs that do not (Ehri, Nunes, Stahl, and Willows 2001).

Still, dissatisfaction with conventional phonics instruction remains. Part of the reason is that phonics instruction is sometimes carried out using drills and worksheets that children find boring. But teachers can go beyond such techniques, helping students to become word detectives who search for patterns in the written language. Students can learn that the patterns they find hold for many words but, in languages like English, not all.

Another problem is that some children have trouble grasping phonics instruction and, even after months of instruction, may be able to read only a few words. Why are some children slow to benefit from phonics instruction? One contributor is a lack of *phonemic awareness*. Children's attention is normally on the meaning of what they hear and say, not on the individual words and sounds. In order to understand how the spellings of words in alphabetic writing systems map onto their spoken forms, children must pay attention to smaller units of sound. For example, a child who cannot analyze the spoken form of *bat* as containing three units of sound, the first of which is the same as the first sound of *boy*, will not understand why the printed form of *bat* contains three letters, the first of which is the same as the first letter of *boy*. A number of tasks have been developed to assess a child's phonemic awareness, ranging from counting phonemes (how many sounds do you hear in "bat"?) to comparing phonemes (do "bat" and "boy" start with the same sound?) to deleting phonemes (what do you get if you take the "b" sound away from "bat"?). Children's performance on such tests is a good predictor of their later reading success, and instruction that is designed to improve phonemic awareness benefits reading in alphabetic systems (Ehri, Nunes, Willows et al. 2001). (Phonemic awareness is less important for nonalphabetic systems, and readers of such systems do not develop the skills in this area that are found among readers of alphabetic systems; see Read et al. 1986.)

To teach phonemic awareness, one can take advantage of the fact that awareness of phonemes is the endpoint of a long developmental process. The process begins with awareness of words and syllables and progresses to units that are smaller than syllables but larger than phonemes, including initial consonant clusters (e.g., the "bl" of "blast") and rimes (e.g., "ast"). Phonemic awareness may be taught gradually, following this progression. For example, children can play games in which they clap once for each syllable in a spoken word before they proceed to the level of phonemes. Phonemic awareness instruction is particularly successful when it is closely integrated with reading instruction, allowing children to learn how the sounds that they are isolating in speech are represented in writing (Ehri, Nunes, Willows et al. 2001).

Another reason why children may have trouble benefiting from phonics instruction may be that, when they first begin to learn to read, they may believe that the links between printed words and concepts are arbitrary (Ehri 2005). For example, young children may think that the color and overall shape of the McDonald's logo, not the letters it contains, are important in allowing it to symbolize what it does. Children must break away from the idea that printed words are holistic symbols in order to learn that their parts (the letters and letter groups) map onto the parts of spoken words (the phonemes) in a systematic fashion.

Yet another stumbling block to conventional phonics instruction involves the teachers rather than the students. Many teachers have little or no opportunity to learn about linguistics and the structure of written language. As a result, they may not provide optimal instruction (Moats 1994). Because teachers are themselves good readers, they tend to think about language in terms of how it is spelled rather than how it is pronounced. They may find it hard to put themselves in the place of a child who does not yet know the conventional writing system. For example, a teacher may think that there is a "short i" sound in the spoken form of *girl* because the spelling of this word contains an *i*. However, the spoken word does not actually contain the same vowel as

bit and it would be misleading to suggest to a child that it does. As Moats (1994: 99) states, "lower level language mastery is as essential for the literacy teacher as anatomy is for the physician. It is our obligation to enable teachers to acquire it."

To summarize, reading instruction that includes explicit and systematic attention to phonics generally works better than instruction that does not. However, there is room for improvement in phonics programs. Improvement can occur by better preparing children to benefit from phonics instruction and by better preparing teachers to teach it.

4 Learning to Spell

One aspect of whole language programs that is attractive to many teachers and parents is the focus on writing. In many whole language classrooms, children write each day in personal journals. Correct spelling is not stressed, with children instead being encouraged to invent spellings for words they do not know. It is assumed that invented spellings like *bo* for *blow*, *grl* for *girl*, and *wet* for *went* will give way to conventional spellings as children learn to read and that systematic instruction in spelling is not necessary. However, research shows that children are less likely to learn words' spellings from the reading of meaningful, connected text than from the study of isolated words. Research further shows that the correlation between reading ability and spelling ability is far from perfect: There are a number of people who are good readers but poor spellers. For most children, learning to spell requires something above and beyond immersion in a sea of print. The benefits of spelling instruction are not confined to spelling itself. Such instruction can also foster reading and phonemic awareness. For example, as children practice spelling consonant clusters like *bl* they learn to analyze these clusters into their component phonemes. Spelling instruction, like reading instruction, requires a teacher who is knowledgeable about children's errors and the reasons behind them. For example, a teacher who knows that the middle part of *girl* is a syllabic "r" sound rather than a "short i" sound followed by a separate "r" will understand why young children frequently misspell this word as *grl*. (See Treiman and Kessler 2014 for further discussion of how children learn to spell.)

5 Dyslexia

Even with good instruction, some children who are developing normally in other respects have great difficulty learning to read and spell. Such children are said to have *dyslexia*. Their comprehension of oral language may be adequate, but their reading comprehension is poor. In most cases, impaired reading comprehension is associated with difficulties at the single-word level, difficulties that also show themselves in poor spelling. The popular view is that children with dyslexia see letters and words backward. As a result, they may misread *was* as *saw* or *day* as *bay*. Similar errors occur in spelling, as when children write *bit* as *dit* or even *tid* (in what is known as *mirror writing*). However, research reveals that such mistakes do not constitute the majority of reading or spelling errors among dyslexics. Moreover, typically developing children sometimes make the same kinds of errors when they are first learning to read and write. Most researchers now believe that, in the great majority of cases, dyslexia does not reflect low-level perceptual problems that extend beyond the reading domain. Dyslexia is, instead, a linguistic problem (Vellutino and Fletcher 2005).

If dyslexia is a linguistic problem, what kind of linguistic problem is it? The most widely accepted hypothesis is that dyslexics have weaknesses in the phonological component of language (Vellutino and Fletcher 2005). Specifically, dyslexics have difficulty becoming aware of the phonemic structure of spoken language and thus have trouble learning about the way in which spellings map onto words' sounds in alphabetic writing systems. Dyslexics' phonological

problems also extend to remembering words and to producing them quickly and accurately. These problems are, in part, genetically based. For example, if one member of a pair of identical twins exhibits reading problems then the other twin has an elevated chance of showing similar problems (Olson 2008).

If dyslexia stems from linguistic weaknesses, particularly weaknesses in the area of phonology, then teaching must attempt to remediate the linguistic problems. Instruction that centers on low-level perceptual skills, such as exercises designed to improve eye tracking or binocular coordination, does not appear to be successful (Vellutino and Fletcher 2005). What is needed, instead, is an intensive reading program that includes a liberal dose of phonics. In one successful program of this kind, as described by Gaskins (1998), children with reading problems spend over four hours a day in literacy activities. These activities are designed to help the children focus on the sounds in spoken words and how these sounds are represented with letters. The children are taught to use their knowledge in reading and writing connected text as well as in reading and spelling individual words.

6 The Effects of Literacy

Does learning to read change people's basic cognitive or linguistic abilities? Some have suggested that literate individuals and societies differ greatly from nonliterate ones, the former being more abstract, more rational, and more skeptical. Although research has not supported these grand claims, it has provided empirical evidence that literacy has certain cognitive consequences (see Stanovich 1993). For example, university students in the United States who read extensively have larger vocabularies and more knowledge about the world than their peers who do little reading in their free time. Opportunities to learn new words tend to arise more often while reading than while watching most sorts of television programs or participating in most everyday conversations.

Learning to read also appears to deepen and alter people's knowledge about language. Awareness of phonemes develops hand in hand with learning to read and write an alphabetic system. Thus, preliterate children and alphabetically illiterate adults tend to do poorly in tasks requiring access to the phonemic structure of language, although they do better on rhyming tasks and syllable-level tasks (Morais and Kolinsky 2005; Read et al. 1986). Another effect of alphabetic literacy is to color people's ideas about the sounds of language. For example, seeing that words like *went* and *elephant* contain an *n* in their spellings, children may come to conceptualize "n" after a vowel as a separate unit of sound rather than as part of the vowel, as they did previously (Treiman, Zukowski, and Richmond-Welty 1995). Pronunciations of words may also change under the influence of spelling, as when people include a "t" in their pronunciation of *often*. If people's ideas about spoken language are indeed influenced by their knowledge of written language, it may be difficult for linguists and psycholinguists to study the structure or processing of spoken language without considering written language.

7 Conclusions and Future Directions

Many modern linguists believe that speech is the primary form of language and that writing is secondary. This view implies that investigations of language and language processing should focus on spoken language and that there is little to be gained from studies of written language. This chapter has presented evidence, to the contrary, that the study of written language processing is interesting and informative in its own right. There are many questions to be answered about how people relate print to speech and about how children can best be taught to do so. This is an area in the study of language that has important real-world applications. Moreover, it appears that written language takes on a life of its own once acquired, influencing the representation and

processing of spoken language. The study of writing and written language processing can no longer be ignored within linguistics.

One emerging area of research, which has not been covered in this chapter for reasons of space, concerns the brain bases of reading. In recent years, researchers have increasingly been studying the areas of the brain that are involved in reading. They are asking how the patterns of brain activation in people with reading problems may differ from the patterns shown by other people and how patterns of brain activity may change as a function of reading instruction. A second emerging area of research is cross-linguistic studies of reading and spelling and their development. As mentioned earlier in this chapter, much research has looked at English and other alphabetic writing systems. It is important to study a broad variety of languages in order to develop theories that are not limited to a particular language or type of writing system, and researchers are currently attempting to do this. A final trend involves the strengthening of links between research and education. The teaching of reading and spelling should be based on evidence about what works, not on fads, and scientific evidence of the sort reviewed in this chapter has an important role to play in education.

SUGGESTIONS FOR FURTHER READING

The books listed below contain authoritative information about various aspects of reading:

Malatesha Joshi, R. and Aaron, P. G. (eds) (2006). *Handbook of Orthography and Literacy*. Mahwah, NJ: Erlbaum.

Grigorenko, E. L. and Naples, A. J. (eds.) (2008). *Single-Word Reading: Behavioral and Biological Perspectives*. New York: Erlbaum.

Pollatsek, A. and Treiman, R. (eds) (2015). *Oxford Handbook of Reading*. New York: Oxford University Press.

RELEVANT JOURNALS AND SOCIETIES

The Society for the Scientific Study of Reading sponsors conferences dealing with the issues discussed in this chapter, particularly those related to reading acquisition and instruction. This society also publishes a journal, *Scientific Studies of Reading*. The Psychonomic Society, an organization for cognitive and experimental psychologists, includes a number of researchers who study reading and word recognition. Journals in the field of experimental psychology, including *Journal of Memory and Language* and *Journal of Experimental Child Psychology*, regularly publish research on reading, and *Reading and Writing: An Interdisciplinary Journal* devotes itself to this topic.

REFERENCES

Ans, B., Carbonnel, S., and Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, 105, 678–723.

Coltheart, M. (2005). Modeling reading: the dual-route approach. In M. J. Snowling and C. Hulme (eds), *Science of Reading: A Handbook* (pp. 6–23). Oxford: Blackwell.

DeFrancis, J. (1989). *Visible Speech: The Diverse Oneness of Writing Systems*. Honolulu: University of Hawaii Press.

Ehri, L. C. (2005). Development of sight word reading: phases and findings. In M. J. Snowling and C. Hulme (eds), *Science of Reading: A Handbook* (pp. 135–54). Oxford: Blackwell.

Ehri, L. C., Nunes, S. R., Stahl, S. A., and Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, 71, 393–447.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., and Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, 36, 250–87.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: true issues and false trails. *Psychological Bulletin*, 123, 71–99.

Gaskins, I. W. (1998). A beginning literacy program for at-risk and delayed readers. In J. Metsala and L. C. Ehri (eds), *Word Recognition in Beginning Literacy* (pp. 209–32). Mahwah, NJ: Erlbaum.

Goodman, K. S. (1967). Reading: a psycholinguistic guessing game. *Journal of the Reading Specialist*, 6, 126–35.

Gough, P. B. (1972). One second of reading. In J. F. Kavanagh and I. G. Mattingly (eds), *Language by Ear and by Eye* (pp. 331–58). Cambridge, MA: MIT Press.

Moats, L. C. (1994). The missing foundation in teacher education: knowledge of the structure of spoken and written language. *Annals of Dyslexia*, 44, 81–102.

Morais, J. and Kolinsky, R. (2005). Literacy and cognitive change. In M. J. Snowling and C. Hulme (eds), *Science of Reading: A Handbook* (pp. 188–203). Oxford: Blackwell.

Olson, R. K. (2008). Genetic and environmental influences on word-reading skills. In E. L. Grigorenko and A. J. Naples (eds), *Single-Word Reading: Behavioral and Biological Perspectives* (pp. 233–53). New York: Erlbaum.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., and Seidenberg, M. S. (2002). How should reading be taught? *Scientific American*, 286, 84–91.

Rayner, K., Pollatsek, A., Ashby, J., and Clifton, C., Jr. (2012). *Psychology of Reading*, 2nd edn. New York: Psychology Press.

Read, C., Zhang, Y.-F., Nie, H.-Y., and Ding, B.-Q. (1986). The ability to manipulate speech sounds depends on knowing alphabetic reading. *Cognition*, 24, 31–44.

Seidenberg, M. S. (2007). Connectionist models of word reading. In M. G. Gaskell (ed.), *Oxford Handbook of Psycholinguistics* (pp. 235–50). Oxford: Oxford University Press.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16, 32–71.

Stanovich, K. E. (1993). Does reading make you smarter? Literacy and the development of verbal intelligence. In H. W. Reese (ed.), *Advances in Child Development and Behavior* (pp. 133–80). San Diego, CA: Academic Press.

Treiman, R. and Kessler, B. (2007). Learning to read. In M. G. Gaskell (ed.), *Oxford Handbook of Psycholinguistics* (pp. 657–66). Oxford: Oxford University Press.

Treiman, R. and Kessler, B. (2014). *How Children Learn to Write Words.* Oxford: Oxford University Press

Treiman, R., Zukowski, A., and Richmond-Welty, E. D. (1995). What happened to the "n" of sink? Children's spellings of final consonant clusters. *Cognition*, 55, 1–38.

Van Orden, G. C. (1987). A ROWS is a ROSE: spelling, sound, and reading. *Memory and Cognition*, 15, 181–98.

Vellutino, F. R. and Fletcher, J. M. (2005). Developmental dyslexia. In M. J. Snowling and C. Hulme (eds), *Science of Reading: A Handbook* (pp. 362–78). Oxford: Blackwell.

Ziegler, J. C., Perry, C., and Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: an extension to German. *European Journal of Cognitive Psychology*, 12, 413–30.

31 Language and Law

ROGER W. SHUY

Although many people commonly think of applied linguistics as language learning, teaching, and testing, a growing number of linguists in recent years have been applying linguistic principles, theory, and research to many other areas of life, such as medical communication, advertising, and the intersection of law and language. This isn't surprising because it's hard to imagine any area of human existence where linguistics cannot be applied. Over the years the legal arena had received attention from anthropologists, psychologists, philosophers, sociologists, and political scientists, but now linguists are applying their field's knowledge to such areas as statutory law and interpretation, voice and authorship identification, jury instructions, the asymmetry of power in courtroom exchanges, lawyer–client communication, police interrogation practices, contract disputes, legal discourse, defamation, trademark infringement, courtroom interpretation and translation, copyright disputes, discrimination, commercial warning messages, and various types of criminal charges such as perjury, bribery, solicitation, money laundering, threatening, and fraud. Virtually all of such cases involve written or spoken language evidence, making linguistic analysis very relevant.

A good case could be made for calling the use of linguistics in the legal arena *applied linguistics*, but, however accurate that term is, in the 1980s *forensic linguistics* began to be used to cover this area and by now it appears to be the common name for the entire area of linguistics and law, no matter what type of involvement. By the 1990s, forensic linguistics had established its own academic organization, the International Association of Forensic Linguistics, its own journal, *The International Journal of Speech, Language and Law* (originally named *Forensic Linguistics*), a rapidly growing number of books and articles, and an increasing number of linguists from all over the world who do this type of work. Some would rather refer to this area as *linguistics and law* while others prefer the even simpler label, *linguistics*. But since *forensic* is commonly described as dealing with the application of scientific knowledge to other areas, that name seems to be accepted.

We can tell that a field is beginning to mature when general introductions start to be published. Forensic linguistics now has several general introductions available. Some focus on parts of the entire field (McMenamin 2002; Olsson 2004) while others have a somewhat broader scope (Gibbons 2003; Coulthard and Johnson 2007). Currently there are also two handbooks covering language and the law (Coulthard and Johnson 2010; Tiersma and Solan 2012).

So what do forensic linguists do? As in linguistics as a whole, linguists who relate language to law have two major avenues in which to do their work: research and practice. Those who choose the avenue of research tend to analyze such things as the language and interpretation of statutes, the written and spoken language of judges and attorneys in the context of trials and hearings, and the language of police interrogations including the subsequent confessions of suspects.

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

Those who choose the avenue of practice leap into the frying pan of actual court cases, analyzing the language evidence of individual criminal and civil court cases, either as consultant-advisors to attorneys or as expert witnesses at trials.

Forensic linguistic practitioners respond to the requests of attorneys to help them with their law cases, including civil cases such as trademark infringement, product liability, deceptive trade practice, copyright infringement, and discrimination relating to age, race, and gender. They are also asked to work on criminal cases of many kinds, such as bribery, money laundering, solicitation to murder, threatening, perjury, and sexual misconduct. Some cases can be either civil or criminal, depending on the specific charges, such as those involving speaker and authorship identification. The following are examples of how linguists practice in civil and criminal cases as well as the linguistic research work they do.

1 Civil Cases

1.1 Trademark infringement

One company holding a trademark may feel that another company's trade name is too much like its own. The more generic or descriptive the name, such as *Raisin Bran* or *Beer Nuts*, the less likely such a name can be protected against use by other companies. The more unique or fanciful the name, such as the coined words *Kodak* or *Xerox*, the more likely such protection will apply.

Trade names that fall between *generic* and *descriptive* at one end of the spectrum and fanciful names at the other end seldom find their way to litigation. Most courtroom battles are fought over the middle categories of *arbitrary* and *suggestive* names. Arbitrary trade names are nonfanciful words in common use but, when used with goods and services, do not suggest nor describe the ingredients, quality, or character of those goods or services. The ordinary meaning of the word is applied to those goods in an arbitrary and nondescriptive sense. McCarthy's *Trademarks and Unfair Competition* (1984) cites a case law ruling about the trade name, *V-8* (vegetable cocktail), that illustrates how courts define arbitrary trade names:

> By repeatedly advertising the fact that its cocktail is made from the combined juices of eight vegetables, the plaintiff has undoubtedly taught the purchasing public that *V-8* on a tin can means such a cocktail. Except for this association, we think, no one could reasonably be expected to know that *V-8* designated a vegetable juice cocktail, or any other particular thing for that matter, unless it be something so described by both shape and number, like an eight cylinder engine, for instance, having cylinder blocks set at an acute angle to each other.
>
> *Standard Brands Inc. v. Smidler (1945 CA2 NY) F2d 34, 66 USPQ 337.*

Trade names found to be arbitrary are also illustrated by McCarthy, who described how case law ruled that *Ivory* (soap) is arbitrary:

> the ordinary meaning is applied to the goods in "a totally arbitrary and non-descriptive sense. Ivory soap is not made of ivory, *Old Crow* (whiskey) is not distilled from old crows, and *Royal* (baking powder) is not used exclusively by royalty." *Stork Restaurant v. Sahati (1948 CA9 Cal 166) F2d 348, 76 USPQ 374.*

Suggestive trade names also are words in common use but they don't describe or even suggest the product's purpose or function. Instead they suggest some quality not indicated by the name itself. The trade names *Camel* (cigarettes), *Shell* (gasoline), and *Arm and Hammer* (baking soda) are commonly cited as illustrations of suggestive trade names. The courts often use "the imagination test" to decide on whether a name is suggestive. This test determines whether or not consumers have to make one or more "mental leaps" to figure out the purpose and function of the product.

In trademark cases the burden of proof rests on the allegedly offended party to show that the junior user's name looks like, sounds like, and means the same as their own. To a linguist, "sounds like" obviously suggests phonological analysis, "looks like" suggests graphemic and semiotic analysis, and "means the same" suggests semantic and pragmatic analysis (Shuy 2002; Butters 2007a, 2008). Another type of trademark case involves trademark dilution, a concept that even trademark lawyers have difficulty defining. To linguists, "dilution" suggests linguistic analysis of language perception and language change over time, among other things (Butters 2007b).

1.2 Product liability

Forensic linguists are also retained in product liability lawsuits (Dumas 1990; Shuy 1990, 2008; Cushing 1994; Tiersma 2002). It may seem surprising that linguistics has anything to do with a complaint that a product has caused injury to a consumer. But consider attorneys with clients who have suffered physical harm alleged to have been caused by inadequate or misleading packaged instructions about how to use the product or by the packaged warnings given about the product itself. A linguist is called upon to analyze the language of the warning or usage messages to determine whether or not they are sufficiently clear, unambiguous, or the most optimally effective way for guiding the user to avoid the dangers in using the product or how closely such warnings relate to the regulations put forth by the relevant regulatory agency.

For example, does a manufacturer satisfy the U.S. Food and Drug Administration's required warning to show that "tampons are associated with Toxic Shock Syndrome" when it labels its package insert warning with the words, "Important Information About Toxic Shock Syndrome (TSS)"? Does "important information about" meet the FDA requirement that the product is "associated with" that danger? Does the fact that the manufacturer's required warning eventually makes such an association in the middle of the warning message satisfy the government regulation that this warning must appear "prominently" and in "terms understandable by the user?" Does the fact that the warning portion of the message is written in complex sentences averaging 19 words per sentence while the "instructions for how to use" section is written in simple sentences averaging 9.4 words per sentence suggest that the writers of the warning really knew how to write more readable sentences but chose not to do so when they wrote the warning section? (Shuy 1990, 2008)

No linguist, of course, can get into the minds of either the warning writers to discover what they intended or the comprehension of consumers who read the message. But linguists, calling on their knowledge of discourse analysis, semantics, syntax, and pragmatics, can determine the extent to which the message was clear and unambiguous and point out the possible meanings that the message presents.

1.3 Contract disputes

Contracts of all types, including insurance policies, provide a fruitful area for linguistic analysis. To most readers, a contract's wording may seem complex, needlessly convoluted, and riddled with jargon. Lawyers may have good reasons for using such language but, if contracts are so carefully constructed, one has to wonder why is it that there are so many lawsuits based on different understandings of their language. Contracts hold interest to linguists because they consist of the speech acts of promises of performance or actions. They also contain conditions, some of which are linguistically well formed and others which are not. The text of contracts offers the linguist an opportunity to determine the range of possible meanings that the contract reveals (Tiersma 2002; Solan 2004; Shuy 2008).

For example, one insurance policy contained a section called "common exclusions" that said, "This policy does not cover any loss covered by or resulting from . . . sickness or disease, except

pyogenic infections which occur through an accidental cut or wound." This policyholder received relatively uncomplicated surgery on his retina to scrape off a macular pucker that had developed. Although his chances of getting a pyogenic infection were miniscule, disaster occurred for him when, after his surgery, he became blind in one eye. Since his surgery was not an "accident," he was sure that his policy covered his condition, but he soon discovered that the insurance company interpreted the above exclusion differently.

The sticking point was the noun phrase, "accidental cut or wound," and a legal battle took place over the use of "or" in this sentence. The company said that "accidental cut or wound," taken as a whole, refers to events that occur outside of the surgical context and therefore this phrase excluded all surgical procedures from liability. The policyholder agreed that "accidental cut" was an event outside the surgical context but that "wound" included surgery. He reasoned that surgeons "cut" when they perform the first step in the surgery, but after the incision is made, it is referred to as a "wound," even by medical personnel. So this battle was over one of those little words in English that are sometimes difficult to define, "or" being one of them.

"Or" has three major meanings in English:

1. alternatives with disjunctive elements, such as "You can have salad *or* soup with your meal," meaning that you are entitled to only one of the alternatives;
2. synonyms for the same elements, such as "Buffalo used to roam free *or* unfettered;"
3. a two-way indefinite choice between separate elements, as in "He works out in the gym two *or* three times a week."

The problem was how to discover which meaning of "or" applied here. The insurer said it meant that "cut" and "wound" were synonyms, both of which referred to accidents. The policyholder said "or" meant it was a disjoint that offered a choice between the two separate nouns.

Support for the policyholder's view was found in 11 of that policy's uses of "or" that appeared in the exceptions section immediately before this "accidental cut or wound" phrase. In all 11 instances "or" was used clearly and unambiguously as a disjunct. This consistent use of "or" as a disjunct enabled and encouraged that policyholder to predict and understand from the discourse context that this last use of "or" was simply one more disjunct that contained two separate and different elements.

In addition, in English constructions containing two scope-bearing elements, the one that comes first has scope over that which follows. This initially negativized expression, "except pyogenic infections which occur through an accidental cut or wound," has two scope-bearing elements, "through," and "accidental." "Through" has scope over the following noun phrase, " an accidental cut or wound." But, as the policyholder argued, the adjective "accidental" does not have scope over "wound" because the expression was negativized by "except." This meant that "wound" is not under the scope of "accidental," because the second scope-bearing element, "accidental" has scope over only "cut," the noun it immediately modifies. The negativizer "except" heads the construction and has scope over what comes after it, conveying the meaning of both "through an accidental cut" and "through a wound." The insurance company's policy writer could have avoided this problem by saying "through either an accidental cut or an accidental wound."

1.4 Speaker identification

The identification of speakers from audio-recordings can be important in both criminal and civil cases. In civil cases the recordings might be made either openly or surreptitiously by someone in a board meeting, for example. When there are many speakers involved, sometimes their identification can be difficult to determine. Voice identification analysis by forensic linguists also can be important in helping identify senders of threats and hate messages. Linguists have been used

by attorneys in matters of voice identification perhaps longer than in most other areas of legal dispute (Baldwin and French 1990; Hollien 1990, 2002; Rose 2002). For example, if an anonymous caller leaves a threatening or hate message on an answering machine and the receiver of this message then takes the recording to an attorney or to a law enforcement agency, a linguist may be called upon to try to help identify the speaker, by comparing the characteristics of that voice with tape with the voices of other potential suspects. If the tapes are of sufficient quality, spectographic analysis is possible. If not, linguists may rely on training and skills in phonetics to make the comparison.

Several problems with such analysis have been posed. For one thing, spectographic analysis is not allowed in some jurisdictions. But, even when it is allowed, the analyst often has the suspects read aloud the words in the original phone message in order to produce comparable words. Some argue that a reading voice is not the same as a talking voice and that taking words out of their normal context distorts the comparison. Some also argue that the readers, having been alerted to their status as suspects, may try to alter their normal speech patterns. There also has been considerable complaint that those who do this work employ a wide array of methods that lack evaluation and validation. On the other hand, some juries tend to be more impressed with an analysis based on electronic equipment rather than on an individual linguist's phonetic judgment, no matter how expert that linguist might be.

In many voice identification cases, linguists use both spectographic and articulatory phonetic expertise to show that suspects were or were not the suspected speakers. For example, in one such case, the vowels of a suspect were shown to be characteristic of an entirely different dialect area than those of the person on a message machine tape, resulting in the suspect's acquittal. In this case, spectographic analysis was used to support the linguistic expert's skills in sociolinguistic variability and auditory phonetics (Labov 1988).

1.5 Authorship

In many law cases written threats and hate messages are the evidence used by law enforcement agencies. These agencies often call on the expertise of psychologists to provide what they call a "psychological profile" of the person who sent the message. The U.S. Federal Bureau of Investigation has a behavioral profiling unit at its Quantico, Virginia academy. It is only recently, however, that this agency has begun to call on linguists to add the dimension of linguistic profiling to their psychological analyses. Calling on their knowledge of language indicators of such things as regional and social dialect, age, gender, education, and occupation, linguists analyze documents for broad clues to the possible identity of the writers (Solan and Tiersma 2004; Shuy 2014). Linguists also use stylistic, syntactic, and corpus linguistic analyses of such documents, usually by comparing the writers' style and syntax with that of other documents written by same suspect and other candidates (McMenamin 1993, 2002; Chaski 2001; Coulthard 2004; Olsson 2004; Kniffka 2007). Such linguistic analysis centers on a writer's habitual language features, over which that writer is thought to have little or no conscious awareness or control, such as patterns of clause embedding, parallel structures, mechanical errors, punctuation patterns, discourse features and organization, and print features such as underlining, bolding, or italicizing, some of which can help identify a writer.

One example of this occurred when a female physician and director of a woman's clinic in Gary, Indiana reported to the police that she had received three threat messages informing her that her clinic would be bombed. Baffled about what to do, the police called in the FBI. When the FBI agents also got nowhere, they called on a linguist to help them. He asked all ten employees of the clinic, including the director, to write a detailed account of everything they did on the day of the most recent bomb threat. He then compared these unsigned documents with the three threat messages. He was interested in how the subjects wrote rather than what they wrote.

The threat messages contained expressions such as, "She will finally the seriousness of the problem recognize," "I will not give warning," "You can be transferred to better position," and "If I address it her." These and other expressions suggested the possible influence of Hindi-Urdu on the English used by the writer. People with such a background might be expected to place the verb at the end of the English sentence and omit articles and pronouns. The linguist also noted other language expressions, such as "I will take the proper course" and "she was in hospital at the time," which suggested a writer who at some point had come under the influence of British English. Even though these threat letters were written in several different handwriting styles, apparently in an effort to disguise the writer's identity, the linguist used the defining characteristics of a Hindi-Urdu and British background to identify the culprit as likely to be a native Pakistani. Oddly enough, only the director's writing sample contained these same features and she was a Pakistani native whose English contained British English features. When confronted with this analysis, the physician/director admitted that she was the actual author of the threats. Now the FBI was baffled again, wondering why the director would write these threat letters to herself. Eventually she explained that her husband had failed his medical license examinations in California. She thought writing these threat messages to herself would create a good excuse for her to close her practice and go back home to help him study for the examinations. Because no crime had been committed, nobody was ever charged, but the case was solved just the same.

It should be pointed out that linguistic profiling and authorship analysis have been used most effectively to narrow down suspect lists rather than to positively identify a specific suspect at trial. This is not to say that positive identification is impossible but, rather that the small amount of data available for the linguist to use, the potential for language variability, the frequent dissimilarity of the genres compared, the difficulty in finding comparable registers for comparison, and lack of comparable size of the texts, can often severely impact any helpful analysis.

1.6 Discrimination

Accusations of discriminatory practice take place primarily in the areas of corporate and business behavior as well as in certain law enforcement practices in which officers of the law or, in some cases even whole police departments, have been accused of discriminatory racial profiling. Beginning in the 1960s, US legislation prohibited racial discrimination and since that time further legislation has prohibited discrimination based on age, employment, handicaps, national origin, religion, language, gender, and sexual orientation.

In one case, for example, certain real estate businesses were suspected of discriminating against potential customers who phoned them to inquire about available homes or apartments to purchase or rent. African American callers were steered to predominantly African American residential areas and away from predominately white areas (Baugh 2007; Shuy 2008). These calls were tape-recorded and analyzed by linguists, who called on their knowledge of research on the phonetics, grammar, and vocabulary of African American Vernacular English to assist prosecutors in their racial discrimination cases against the real estate companies.

Age discrimination, sometimes practiced by corporations as they try to get older employees to leave their companies, is also a fruitful area for forensic linguists. In one such case, a mid-level executive in his fifties was terminated despite his very positive periodic company evaluations. Through time-consuming analysis of hundreds of media articles about that company and its top executives as well the published speeches of the president of that company, a linguist isolated quotations of stereotyped age discrimination by that company in its hiring and firing processes. Even though no specific discrimination evidence was found relating to the specific employee who brought the lawsuit against the company, these documents provided sufficient basis to show that the company's policy evidenced a definite positive bias in favor of younger employees and a considerable number of negative stereotypes of older ones (Shuy 2008).

1.7 Copyright infringement and plagiarism

Although copyright laws are neither simple nor precise, the general idea is to protect the original writer's ideas, concepts, principles, procedures, and methods of operation from being borrowed, copied, and used by others as their own. The laws consist of a somewhat vague rule of reason that takes into consideration such things as the extent of the borrower's use, the purpose of that use, and the effect of that use on the market of the copyrighted work. Most copyright cases involve the legal concepts of proportion, substantiality, originality, and substantial similarity, all of which are fairly difficult to define. Some borrowings fall under the complex doctrine of fair use, which allows a second writer to use the original for commentary, criticism, scholarship, or parody. But even the extent to which this is done can be hard to pin down and define. So far at least, there have been relatively few reports by linguists about their work on copyright infringement cases. One might predict more linguistic activity in this area, since it requires the comparison of texts with each other, particularly in disputes about what constitutes substantial similarity and how the expression of an idea is defined and executed.

A potentially useful linguistic analysis in copyright infringement cases would compare the way the borrower uses the sounds (represented by orthography), affixes, words, phrases, clauses, sentences, speech acts, and discourse structure of the copyrighted material. Orthographic representations such as punctuation practices sometimes may also prove diagnostic. On the whole, however, sounds and affixes are not the most fruitful units to measure, because their frequencies and inventories are so large and mutually available. Other than when there are obvious lexical or affix substitutions, individual words often also are not the best indicators of borrowing, except as they occur in the larger context of phrases, clauses, or sentences. For example, if the original of two texts about how to save on a car's gasoline consumption said, "Don't ride the clutch to keep your car standing still on an incline" and the later published text said, "Don't ride the clutch to keep your car at a standstill on hills," a reasonably good case could be made for borrowing that occurs at the levels of both syntax and lexicon (Shuy 2008).

Some borrowers do little more than to use lexical substitutions, such as "harm" for "damage," or grammatical variations that substitute a singular for a plural noun. But it would seem obvious that other levels beside syntax and lexicon could be used in analyzing copyright evidence. From discourse analysis, for example, one might consider the comparison of topic sequencing as a possibly useful measure. And from speech act theory, one might compare the speech acts used in both texts, as well the patterned sequences of those speech acts for comparison of similarities or differences.

Linguists have been more active in copyright's sister area, plagiarism, particularly as it occurs in the academic context. Plagiarism has characteristics that are very similar to those found in copyright disputes (Olsson 2004; Coulthard and Johnson 2007). Unlike some European jurisdictions, in the USA plagiarism is more a moral issue than a legal one. As a result, plagiarists in the academic context are punished outside of the legal system.

1.8 Defamation

Since people sometimes use language to defame each other, laws have been created to regulate libel (written defamation) and slander (spoken defamation), although it wasn't always this way. In previous centuries, angry men who felt insulted and defamed settled their differences in bloody duels. Today's defamation laws have replaced swords and pistols with a more civilized approach, one that requires the plaintiff to show that the defamatory expression has been published, that it has the force of an untrue accusation, and that it is not couched as a mere opinion (Tiersma 1987). In most cases, if these things can be proved at trial, a plaintiff can receive a monetary reward from the defendant.

Although anyone can sue anyone else for defamation, most defamation lawsuits are brought against the media, whose reporters (ever since the *New York Times v. Sullivan* case in 1964) have been somewhat protected against libel suits unless the plaintiff can show that the defendant's words demonstrated clear evidence of "actual malice." Linguists are skilled in language use, suggesting that definitions of actual malice might fall within their area of expertise.

Forensic linguists are trained to analyze semantics, grammatical referencing, speech acts, conveyed meaning, language clues to intentions, discourse structure and framing, and the characteristics of malicious language. Defamation is another of the areas in which there has not been much literature describing the past experience of forensic linguists, but this paucity of activity is not likely to continue.

One example of a defamation case that made use of a forensic linguist was a case brought by a woman against a local television station and a deputy sheriff (Shuy 2010). She claimed that this station had defamed her in a three-part television news series in which she believed she was being accused as the person who murdered her husband. The police were baffled by that case and did not have adequate evidence to indict her for this murder, although, as the deputy sheriff stated, she was the number one suspect. More than once in these programs, the deputy sheriff stated that the woman was "the only suspect." Note his use of both "the" and "only" here, rather than "a suspect." He also said, "We don't have enough evidence to prove that she did it beyond a reasonable doubt but I think we have a lot of the reasons why." His "why" is an unfinished expression but it strongly suggests, "why she did it."

More damaging to the stations' defense, however, was the deputy sheriff's statement on the third program: "The suspect walked directly into that house, up the stairs, into the bedroom, and shot the man right between the eyes while he was sleeping." Earlier he had said the woman was the "only" suspect. Now he said that the suspect (there was only one) shot and killed the man. Having the opinion that the woman was the only suspect is one thing but explicitly indicating that this "only" suspect was the one who committed the murder constitutes the representation of a fact, not an opinion. It was a specific accusation that she was the murderer. Among other things, the job of the linguist in this case was to clarify discourse referencing such as this.

2 Criminal Cases

So far we've shown how forensic linguists are used in various types of civil cases, including cases that involve author or speaker identification. Many criminal cases are based on written language evidence as well, including those with documentary records relating to such crimes as money laundering, price fixing, proprietary theft, solicitation to murder, perjury, making false statements, and fraud. Some criminal cases, however, involve electronically recorded spoken language. Before the advent of tape-recording the suspected targets, law enforcement had to rely on circumstantial evidence and eyewitness testimony because it was not possible to capture a crime as it actually took place or was about to take place. Technological advances have enabled law enforcement to do covert tape-recording, which has been able to create stronger inculpatory evidence.

The technological advances that continue to be made have opened the door to a wide range of criminal case analyses in the past half-century. Since the mid-1970s, law enforcement agencies have used tape recorders to capture solicitation to murder in progress. In perjury cases the court provides both written and spoken records of interviews and testimony that are available to be analyzed by linguists (Shuy 2011). In other criminal cases suspects are recorded with court-authorized wiretaps placed in such a way that none of the speakers are aware of being recorded. In cases in which one of the speakers deliberately tapes a suspect, or when undercover agents wear body microphones to engage suspects in face-to-face conversations, US law does not require law enforcement to first obtain court authorization in order to make surreptitious body mike

recordings. American laws regarding surreptitious telephone taping vary between jurisdictions, some requiring the consent of only one of the parties (obviously, the one doing the taping) while other jurisdictions prohibit the practice altogether unless both parties knowingly consent.

The linguist can be brought into such cases either by the prosecution or the defense. If the law enforcement agency is concerned about the adequacy of the language evidence that they have gathered, they may call on a linguist to make their own transcripts or to correct already existing transcripts of the conversations. Linguists also analyze the tape-recordings to determine whether or not the agents' representations of illegality have been made clearly and unambiguously and whether or not the target has felicitously suggested or agreed to the illegal act. When the defense attorney calls on the linguist, the same tasks are central.

After the electronically recorded evidence is gathered and the suspect is indicted, copies of the tapes must be turned over to the defense as part of what is called "discovery." As soon as it is reasonably possible, the prosecution is then required to produce written transcripts of the recordings and turn them over to the defense. In most cases, this exchange of evidence occurs well in advance of the trial, so that the defense can have adequate time to prepare.

In transcript preparation, linguists use the tools of their trade, depending on the specific task. Preparing a jury-ready transcript requires good hearing ability, access to good listening (or viewing) equipment, and knowledge of language variation, syntax, semantics, and phonology. It may not be surprising to learn how difficult it is to produce an accurate, jury-ready transcript of conversations that are recorded in noisy restaurants or with other external interfering noise. Legal battles sometimes ensue about the often important differences found between competing transcripts prepared by the prosecution and those made by the defense.

At trial, a jury will listen to the tapes but it is also commonplace for the court to provide juries with transcripts in order to make their task simpler. The tape is the evidence and the transcript is considered only an aid to the jury. But since people tend to remember what they see much better than what they hear, accurate transcripts are very important for this reason alone. In any case, conversations recorded by body microphones are often very difficult to hear. They are often recorded in restaurants, bars, with street traffic noises, and under conditions that do not promote easy hearing for later listeners. For example, if the government transcript represents a person saying "I'd take it today" when the words on the tape are actually "I wouldn't take it today," serious jury misperceptions can occur (Shuy 1993).

It is not always clear exactly how government transcripts are produced but it appears that prosecutors usually employ an office secretary or a court reporting service to produce them, after which the participating undercover agent is given the opportunity to review the transcript and correct any perceived errors and omissions. When the defense makes a transcript, the same general procedure obtains, except that the task of reviewing and correcting is done by defendants and their attorneys. The objectivity of such reviewing by either side is suspect since the schemas of participants sometimes cause them to think they hear something that is not actually on the tape. An outsider to the case, such as a linguist, does not (or should not) have such schemas or biases.

The content of the tape-recorded evidence suggests the use of the other tools of the linguist, including syntax, morphology, semantics, pragmatics, dialectology, and discourse analysis. The presentation of such analyses can give linguistic laypersons (such as juries) scientific reasons for their perceptions, opinions, or feelings that might otherwise be arbitrary or ungrounded. Likewise, such analyses can help laypersons (such as juries) to see patterns of language use that are visible only through the help of expert linguists. Just as the training and expertise of medical professionals enables them to describe and define the meaningful content of an X-ray, so linguistic experts can describe and define the meaningful content of tape-recorded conversations.

Grammatical referencing is often unclear in everyday language and unless the reference that the prosecution believes to be critical is actually clear and unambiguous, the prosecution's case may fail. In one criminal case in which a defendant was charged with agreeing to purchase

narcotics for resale in order to make enough money to save his company from bankruptcy, the following utterance was considered central to the prosecution's case (Shuy 1993):

Undercover agent: What do you think about investment?
Target: I think investment would be a good thing.

The law enforcement officer thought that the target had thereby agreed to invest his remaining assets into the illegal drug scheme proposed by that agent. In thinking this, however, the agent overlooked the fact that during his five months of tape recording this target, he had clearly failed to elicit anything illegal. Up to that point, in fact, the agent, posing as a legitimate banker, had continuously offered the target two potential avenues of action, both completely legal. One was to obtain a bank loan for the target and the other was to help him find some investors who might purchase stock in his company. In the sixth month of these conversations, the "banker" now changed his story, saying that he couldn't get the target a loan, but that he was deeply involved in a cocaine business on the side and if the target would use his remaining company resources to purchase drugs, the agent could sell the cocaine quickly and the target would reap a large, quick profit and thereby avoid bankruptcy. Stunned by this revelation, the target did not offer a blanket yes or no, because he didn't want to erase the remaining potential of getting stock investors. He believed that saying no to the drug proposition would kill that last remaining possibility. In fact, the target was no clearer than the agent about what "investment" meant here. Neither man had specified the defined referent to which "investment" could be associated, although the target's continuing contextual meaning of "investment" over the previous five months was made perfectly clear by the linguist.

In cases of ambiguous referencing such as this, prosecutors normally defer making an indictment until they have elicited unambiguous statements from the target. In this particular case, the government's intelligence analysis proved faulty. An indictment was made on the assumption that the target had actually agreed to invest in the drug scheme. At trial, the defendant was acquitted, thanks at least partially to the assistance of the linguist in this case.

Discourse analysis is another important tool used by forensic linguists, especially in cases involving tape-recorded conversations. Topic introduction and recycling, for example, provide very useful clues to the agenda or intentions of a speaker. As noted earlier, the linguist cannot know for sure what the speakers' intentions really are, but a careful examination of the topics they bring up offers a useful indication of what they are thinking about, what is foremost in their minds, and, perhaps even more important, what is not on their minds.

Likewise, a careful analysis of the responses that given persons make to the topics introduced by others offers a similar clue to their agendas and intentions. If speakers either agree or disagree clearly, the evidence is pretty clear. But if they change the subject or say nothing at all about it, this offers an indication that this topic was not to their liking, indicates that that they are not interested in it, or that they were politely side-stepping it. Such responses offer a reasonable clue to their intentions.

When the target responds to another person's topics with only feedback markers such as "uh-huh," law enforcement agencies often become confused. Many times, the prosecution tends to consider "uh-huh" to be an agreement or understanding when, in fact, it is only a feedback marker used to indicate that the listener is still listening but not necessarily agreeing, that he or she doesn't really understand the gist of the topic but will hear it out anyway, or that he or she is not really listening at all but is just making polite social noises. It is common for recorded conversation used as evidence in criminal cases to contain examples of feedback markers that the prosecution erroneously attributes to understanding and agreement.

In one recent example, an undercover agent tried to elicit agreement from his boss that a commission his company paid to a consultant violated government regulations. The situation was complex, since this consultant was at the same time receiving perfectly legal commissions

from another unrelated source. The portion of their conversation that the prosecution considered inculpatory was the following:

> Undercover Agent: Ari's calling me every day. . . He's worried, I guess, about his payment, his commission.
>
> Target: Uh-huh.

The target's "uh-huh" can indicate his agreement that Ari is worried about getting his commission but, since there is no specification about what this commission is for or who it is from, there is nothing here that would indicate that he agrees that this commission is from the target's company. Nor is there anything in the preceding or following discourse indicating that the target is even interested in Ari's problems about getting this commission, wherever it was from. Further complicating this exchange is the contextual fact that the undercover agent had made himself a general nuisance by spending far too long updating his very busy boss with what he considered somewhat trivial information. It is even possible that the target's "uh-huh" was the polite noise-making type that had nothing to do with his understanding or agreement.

Armed with the linguist's analysis of this passage, the attorney successfully presented it at trial by himself. Perhaps the best help that a linguist can give an attorney is to make the analysis so clear that it does not even need to be presented by an expert witness. Good attorneys can use this analysis in their cross-examinations and closing arguments.

2.1 Police interrogation and eliciting confessions

In criminal cases, linguists apply their knowledge of speech acts to significant issues involving offers, promises, denials, and agreements found in tape-recorded evidence (Shuy 1993). They also use the tools of linguistics when they examine confessions elicited during police interrogations in an effort to make clear what, if anything, was actually confessed and whether or not that alleged confession was tainted by the manner in which it was elicited (Shuy 1998a; Heydon 2005; Rock 2007).

Law enforcement agencies are gradually beginning to record such confessions electronically in order to guard against accusations of undue police pressure or improper promises. Such recordings make it possible for forensic linguists to apply their skills for either the prosecution or the defense. But even when confessions are electronically recorded, some law enforcement agencies tape only the final part of the interrogations – the part that captures the eventual confession statement. Such practice can convince juries but it tells them nothing about the manner in which that confession was obtained.

Some argue that it would be better for law enforcement to videotape the entire interrogation so that later viewers, such as juries, could be assured that no coercion took place and to be able to determine whether the interrogators put words into the suspects' mouths. False confessions have been thought to fall into the academic territory of psychologists (Leo 2008), but the confession event is a rich source for linguistic analysis as well.

3 Research on the Language of Law

3.1 Jury instructions

To this point, the focus here has been on *practicing* forensic linguistics in actual law cases. In addition to the types of law cases noted above, there are many other areas where linguistics associates closely with law. Research represents one part to this field. For example, considerable research has been done on the court's problem of producing clear jury instructions (Charrow and Charrow 1979; Tiersma 1999, 2001; Dumas 2000; Marder 2009). Before the jury

retires to deliberate, the judge reads a list of instructions that are intended to guide them in arriving at their verdict. Linguists have found that such instructions are sometimes incomprehensible to the jurors. Furthermore, when a deliberating jury sends out a note requesting clarification, the judge often simply repeats the same instructions. Judges walk a fine line here, trying to translate the language of law into the language that laypersons can comprehend without creating a reversible error that could lead to retrial.

3.2 Statutes and statutory interpretation

Some linguists are also lawyers, giving them that expertise to analyze statutes, laws, and their interpretations. Linguistic analysis of statutes and legal rules, interpretation, and procedures tends to be done by law professors who have strong linguistic backgrounds (Ainsworth 2006). Solan's book, *The Language of Statutes* (2010) provides many examples of the problems posed by the language of statutes, including such topics as the role of intentionality, battles over ordinary versus definitional meaning, the effect of language change on statutory interpretation, the linguistic coherence of statutes and procedures, and the disputes over familiar legal terms such as "knowingly," "intentionally," "reasonable person," "public official," "harm," and others. Important work in this area also has been carried out by Ainsworth (1993), Solan (1999, 2004, 2005, 2009), Hutton (2009), and Tiersma (1999, 2010).

Specific types of law cases that intersect with linguistics also have received considerable attention from law scholars who are also linguists. Tiersma has described linguistics issues in perjury (1990), defamation (1987), and product liability (2002). Solan has written about linguistic aspects of the language of contracts (2001, 2004).

3.3 Bureaucratic language

Examples of the problems that government agencies have when they try to communicate statutes to the public, often resulting in what has been called "bureaucratic language," are described by Shuy (1998b). Such language may be appropriate and necessary within law but, like all specialized language, it faces serious comprehensibility problems when it is used to communicate outside its own province, especially to laypersons. Yet this is what bureaucracies need to do as effectively as possible. Linguists can be very helpful in this.

3.4 Language of the courtroom

Some linguists focus their research on the ways language is used by lawyers, judges, and witnesses in the courtroom (Kurzon 1986; Levi and Walker 1990; Gibbons 1994, 2003; Tiersma 1999; Gaines 2000; Solan and Tiersma 2005), especially on the ways lawyers and judges use it during legal procedures (O'Barr 1982; Conley and O'Barr 1990, 1998; Solan 1993; Tiersma 1993; Matoesian 1993, 2001; Phillips 1985, 1998; Stygall 1994; Cotterill 2003; Heffer 2005; Schane 2006; Rock 2007).

Some linguists conduct research concerning the asymmetry of power in courtrooms and police stations (O'Barr 1982; Lakoff 1990; Ainsworth 1993; Cotterill 2003; Shuy 2005). At trial as well as during police interrogations, language power clearly resides on the side of judges, attorneys, and law enforcement. The testimony of witnesses and suspects is controlled by a rigid question–answer format that enables attorneys, judges, and police interrogators to structure and control the sequence and content of what is said. However useful or necessary this may be for trial and police interviewing efficiency, it runs counter to laypersons' natural, normal ways of using language. It prevents them from introducing their own topics or telling their own stories in their own familiar ways. Consequently, they often run a serious risk of not communicating what they want to get said. Since they are so linguistically vulnerable, they can be accused of being

untruthful by their manner of speech, fettered as it is by courtroom and police station constraints (Conley and O'Barr 1998). One of the most frustrating experiences reported by even expert witnesses is that they are forced to answer "yes" or "no" to questions for which neither answer is complete, appropriate, or adequate. Such experiences are not unlike trying to communicate in a different language and culture.

It is not uncommon that defendants, plaintiffs, or witnesses in court are not native speakers of English. This poses serious problems for them as well as for attorneys and judges. Berk-Seligson deals with the difficult problems of Hispanics during trials (1990, 2002) and also during police interrogations (2009). Eades addresses the serious problems of Australian Aborigines when they are interviewed by the police and when they appear in the English-speaking courts (2008, 2010).

Further complications come from the linguistic naivete of some courts in these matters. For example, when undercover tape-recordings are made in languages such as Mandarin or Spanish, US courts frequently use a translator to provide written English translations of the conversations and then present those translations as the evidence. Missing is the essential middle step, that of first providing a transcript of the spoken conversations in the first language upon which the translation is based. The same errors in transcript preparation noted for transcripts of spoken English can easily occur in non-English languages, but the former can be checked against the taped evidence. Without the necessary intermediate transcript in the first language, any possible dispute about the accuracy of the translation is seriously limited. Going straight to the translation without first providing a transcript of the non-English speech may be time and cost efficient for the courts, but it is linguistically naive, if not dangerous for the goal of justice. Linguists have been trying to make this point to the courts in recent years.

Interpreting for the Deaf and hearing impaired provides even further complications in the courtroom since American Sign Language does not easily translate into neat and conventional English sentences and expressions (Lucas 2002). In recent years, linguists have begun to research this issue and many courts are now providing ASL interpreters when they deem it necessary. The US Federal Court Interpreters Act of 1978 (and its trickle-down effect on state and municipal courts) was designed to avoid denying the constitutional rights of the non-native English speakers, including the Deaf and hearing impaired, in the American court system. However worthwhile and necessary such legislation is, a number of problems remain (Berk-Seligson 1990, 2009). How can we best recruit and train such interpreters? What standards should apply to translators and interpreters? How can we provide such services for all of the many possible languages for which there is need? How do we know the extent to which non-English speakers have mastered enough English to be able to use it effectively to help their lawyers defend them? How do we deal with residual cultural differences of non-native speakers?

4 The Future of Linguistics and the Law

During the relatively brief history of forensic linguistics, we have seen a number of developments, including the advent of new technologies, such as electronic recordings, e-mails, and text messages used as evidence. Each new development calls for linguists to find new ways of adapting them to their use of linguistic principles and analytical procedures. But these principles and analytical procedures remain fairly constant. Forensic linguistics is, after all, primarily the study and use of linguistic tools in the legal context and even these tools continue to develop rapidly. When linguists first began this work, speech acts, pragmatics, and discourse analysis were relatively unheard of. Today, along with newer developments in the tools of phonology, syntax, semantics, and sociolinguistics, the potential for using linguists in law cases is even richer.

In terms of research on interpretation of statutes, the language of the courtroom, and the language of judges, lawyers, and witnesses, we also see a wide open area for further linguistic study. In terms of forensic linguistic practitioners, especially in the area of authorship identification,

we need to learn more about how to deal with small evidence samples, to learn which linguistic features best identify authors, and to discover how many occurrences of any linguistic feature are enough to make such identification possible. And we are only scratching the surface about how to deal with bilingual courtroom issues.

Although the field is rapidly expanding, both researchers in language and law and forensic linguistics practitioners need to learn to communicate more effectively with the field of law so that our knowledge, theory, research findings, and skills will be more available to lawyers and judges, who with some justification already pride themselves on their knowledge of how language works.

Practitioners also will need to do a better job of self-regulation, making clear what constitutes an expert in this field, so that the marginally qualified do not sully the reputation and potential usefulness of those who are true linguistic experts. Practitioners also need to determine the best way to prepare future forensic linguists without setting up still other separate academic departments that might remove training further away from their essential core field of linguistics.

In the United States, more and more courts permit linguists to testify as experts at trial (Wallace 1986; Solan and Tiersma 2005). In those few criminal cases in which judges have ruled that linguists' testimony would not be permitted, the reason given was often based on representations made by attorneys and judges about what linguists *might* say rather than on what the linguists would have said had they been allowed to make their own representations of their proposed testimony. Linguistic experts are virtually never rejected based on the scientific reputation and respectability of the field of linguistics, which by now appears to be established in the courts. But whether the linguist testifies or not, it is clear that more and more attorneys and government agencies are calling on forensic linguists to assist them in analyzing the spoken and written language that frames the evidence in both civil and criminal law suits. It is also the case that more and more universities are offering applied linguistics courses on topics of language and the law.

Since expression of the content as well as the presentation and delivery of law is largely dependent on language, the future of linguistics appears to be very promising in the legal arena.

SUGGESTIONS FOR FURTHER READING

Because there is no established core of readings in the field of linguistics and law, the following brief list of books is only a suggestion for those who wish to get a basic understanding of the field.

Andrus, J. (2015). *Entextualizing Domestic Violence.* New York: Oxford University Press.

Angermeyer, P. (2015). *Speak English or What?: Code-switching and Interpreter Use in New York City Courts.* New York: Oxford University Press.

Conley, R. (2016). *Confronting the Death Penalty.* New York: Oxford University Press.

Cotterill, J. (ed.) (2007). *The Language of Sexual Crime.* Basingstoke: Palgrave Macmillan.

Dupret, B., Lynch, M., and Berard, T. (eds) (2015). *Law at Work.* New York: Oxford University Press.

Durant, A. and Leung, J. (2015). *Language and Law: A Resource Book for Students.* London: Routledge.

Ehrlich, S. (2001). *Representing Rape: Language and Sexual Consent.* London: Routledge.

Ehrlich, S., Eades, D., and Ainsworth, J. (eds) (2016). *Discursive Constructions of Consent in the Legal Process.* New York: Oxford University Press.

Gaines, P. (2016). *From Truth to Technique at Trial.* New York: Oxford University Press.

Gibbons, J. and Turrell, T. (eds) (2008). *Dimensions in Forensic Linguistics.* Philadelphia, PA: John Benjamins.

Gruber, M. C. (2014). *"I'm Sorry for What I've Done."* New York: Oxford University Press.

Kjær, A. L. and Adamo, S. (eds) (2011). *Linguistic Diversity and European Democracy.* Farnham: Ashgate.

Kniffka, H. (1996). *Recent Developments in Forensic Linguistics.* Frankfurt: Peter Lang.

Kopaczyk, J. (2013). *The Legal Language of Scottish Burghs.* New York: Oxford University Press.

Mellinkoff, D. (1993). *The Language of Law.* Eugene, OR: Resource.

Mertz, E., Ford, W., and Matoesian, G. (eds) (2016). *Translating the Social World for Law.* New York: Oxford University Press.
Rosulek, L. (2015). *Dueling Discourses.* New York: Oxford University Press.
Shuy, R. W. (2012). *The Language of Sexual Misconduct Cases.* New York: Oxford University Press.
Shuy, R. W. (2013). *The Language of Bribery Cases.* New York: Oxford University Press.
Solan, L., Ainsworth, J., and Shuy, R. (eds) (2015). *Speaking of Language and Law.* New York: Oxford University Press.
Wagner, A. and Cheng, L. (eds) (2011). *Exploring Courtroom Discourse.* Farnham: Ashgate.

RELEVANT JOURNALS AND SOCIETIES

Journals
American Speech
Language
Law and Society
The International Journal of Speech, Language and the Law (formerly *Forensic Linguistics*)

Societies
The American Dialect Society
The International Association of Forensic Linguists
The International Association for Forensic Phonetics and Acoustics
The Linguistic Society of America
The Law and Society Association

REFERENCES

Ainsworth, J. (1993). In a different register: the pragmatics of powerlessness in police interrogation. *The Yale Law Journal,* 103, 2, 259–322.
Ainsworth, J. (2006). Linguistics as a knowledge domain in the law. *Drake Law Review,* 54, 3, 651–69.
Baldwin, J. and French, P. (1990). *Forensic Phonetics*. London: Pinter.
Baugh, J. (2007). Attitudes toward variation and ear-witness testimony. In R. Bayley and C. Lucas (eds), *Sociolinguistic Variation: Theories, Methods, and Applications* (pp. 338–48). Cambridge: Cambridge University Press.
Berk-Seligson, S. (1990). *The Bilingual Courtroom: Court Interpreters in the Judicial Process.* Chicago, IL: University of Chicago Press.
Berk-Seligson, S. (2002). The Miranda warnings and linguistic coercion: the role of footing in the interrogation of a limited-English-speaking murder suspect. In Janet Cotterill (ed.), *Language in the Legal Process* (pp. 127–43). Basingstoke: Palgrave Macmillan.
Berk-Seligson, S. (2009). *Coerced Confessions: The Discourse of Bilingual Police Interrogations.* Berlin: Mouton de Gruyter.
Butters, R. (2007). Changing linguistic issues in U.S. trademark litigation. In T. Turell, J. Cicres, and M. Spassova (eds), *Proceedings of the Second European IAFFL Conference on Forensic Linguistics/Language and the Law.* Barcelona: Publications de I'IULA, no. 19.
Butters, R. (2007). A linguistic look at trademark dilution. *Santa Clara Computer and High Technology Law Journal*, 24, 101–13.
Butters, R. (2008). Trademarks and other proprietary terms. In J. Gibbons and T. Turell (eds), *Dimensions of Forensic Linguistics* (pp. 231–47). Amsterdam: John Benjamins.
Charrow, V. R. and Charrow, R. P. (1979). Making legal language understandable: a psycholinguistic study of jury instructions. *Columbia Law Review*, 79, 1306–74.
Chaski, C. (2001). Empirical evaluations of language-based author identification techniques. *Forensic Linguistics*, 8, 1, 1–65.
Conley, J. and O'Barr, W. (1990). *Rules Versus Relationships: The Ethnography of Legal Discourse.* Chicago, IL: University of Chicago Press.
Conley, J. and O'Barr, W. (1998). *Just Words: Law, Language and Power.* Chicago, IL: University of Chicago Press.
Cotterill, J. (2003). *Language and Power in Court*. Basingstoke: Palgrave Macmillan.
Coulthard, M. (2004). Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*, 25, 4, 431–47.
Coulthard, M. and Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence.* London: Routledge.
Coulthard, M. and Johnson, A. (2010). *The Routledge Handbook of Forensic Linguistics.* Oxford: Routledge.

Cushing, S. (1994). *Fatal Words: Communication Clashes and Aircraft Crashes.* Chicago, IL: University of Chicago Press.

Dumas, B. (1990). An analysis of the adequacy of federally mandated cigarette package warnings. In J. N. Levi and A. G. Walker (eds), *Language in the Judicial Process* (pp. 309–52). New York: Plenum.

Dumas, B. (2000). Jury trials: lay jurors, pattern jury instructions, and comprehension issues. *Tennessee Law Review*, 67, 3, 701–42.

Eades, D. (2008). *Courtroom Talk and Neocolonial Control.* Berlin: Mouton de Gruyter.

Eades, D. (2010). *Sociolinguistics and the Legal Process.* Bristol: Multilingual Matters.

Gaines, P. (2000). Communicating with juries: episodic representational structure in cross-examination. *Tennessee Law Review*, 67, 3, 599–625.

Gibbons, J. (ed.) (1994). *Language and the Law.* London: Longmans.

Gibbons, J. (2003). *Forensic Linguistics: An Introduction to Language in the Justice System.* Oxford: Blackwell.

Heffer, C. (2005). *The Language of Jury Trials*. Basingstoke: Palgrave Macmillan.

Heydon, G. (2005). *The Language of Police Interviewing.* Basingstoke: Palgrave Macmillan.

Hollien, H. (1990). *The Acoustics of Crime*. New York: Plenum.

Hollien, H. (2002). *Forensic Voice Identification*. New York: Academic.

Hutton, C. (2009). *Language, Meaning and the Law:* Edinburgh: Edinburgh University Press.

Kniffka, H. (2007). *Working in Language and Law*. London: Continuum.

Kurzon, D. (1986). *It Is Hereby Performed: Explorations in Legal Speech Acts.* Amsterdam: John Benjamins.

Labov, W. (1988). The judicial testing of linguistic theory. In D. Tannen (ed.), *Linguistics in Context: Connecting Observation and Understanding* (pp. 159–82). Norwood, NJ: Ablex.

Lakoff, R. (1990). *Talking Power: The Politics of Language*. New York: Basic Books.

Leo, R. A. (2008). *Police Interrogations and American Justice.* Cambridge, MA: Harvard University Press.

Levi, J. N. and A. G. Walker (eds) (1990). *Language in the Judicial Process.* New York: Plenum.

Lucas, C. (ed.) (2002). *Language and the Law in Deaf Communities.* Washington, DC: Gallaudet University Press.

Marder, N. (2009). Jury reform: the impossible dream? *Tennessee Journal of Law and Policy*, 5, 149–83.

Matoesian, G. (1993). *Reproducing Rape: Domination through Talk it the Courtroom*. Chicago, IL: University of Chicago Press.

Matoesian, G. (2001). *Law and the Language of Identity: Discourse in the William Kennedy Smith Rape Trial.* New York: Oxford University Press.

McCarthy, J. T. (1984). *Trademarks and Unfair Competition*, Vol. 1. San Francisco, CA: Bancroft-Whitney.

McMenamin, G. R. (1993). *Forensic Stylistics*. Amsterdam: Elsevier.

McMenamin, G. R. (2002). *Forensic Linguistics: Advances in Forensic Stylistics.* Boca Raton, FL: CRC.

O'Barr, W. (1982). *Linguistic Evidence: Language, Power and Strategy in the Courtroom.* New York: Academic.

Olsson, J. (2004). *Forensic Linguistics*. London: Continuum.

Phillips, S. U. (1985). Strategies for clarification in judges' use of language: from the written to spoken. *Discourse Processes*, 8, 421–36.

Phillips, S. U. (1998). *Ideology in the Language of Judges*. New York: Oxford University Press.

Rock, F. (2007). *Communicating Rights: The Language of Arrest and Detention.* Basingstoke: Palgrave Macmillan.

Rose, J. P. (2002). *Forensic Speaker Identification*. London: Taylor & Francis.

Schane, S. (2006). *Language and the Law.* London: Continuum.

Shuy, R. W. (1990). Warning labels: language, law and comprehensibility. *American Speech*, 65, 291–303.

Shuy, R. W. (1993). *Language Crimes*. Oxford: Blackwell.

Shuy, R. W. (1998). *The Language of Confession, Interrogation and Deception.* Thousand Oaks, CA: Sage.

Shuy, R. W. (1998). *Bureaucratic Language in Government and Business.* Washington, DC: Georgetown University Press.

Shuy, R. W. (2002). *Linguistic Battles in Trademark Disputes.* Basingstoke: Palgrave Macmillan.

Shuy, R. W. (2005). *Creating Language Crimes*. New York: Oxford University Press.

Shuy, R. W. (2008). *Fighting over Words: Language and Civil Law Cases.* New York: Oxford University Press.

Shuy, R. W. (2010). *The Language of Defamation Cases.* New York: Oxford University Press.

Shuy, R. W. (2011). *The Language of Perjury Cases.* New York: Oxford University Press.

Shuy, R. W. (2014). *The Language of Murder Cases.* New York: Oxford University Press.

Solan, L. M. (1993). *The Language of Judges.* Chicago, IL: University Press of Chicago.

Solan, L. M. (1999). Refocusing the burden of proof in criminal cases: some doubt about reasonable doubt. *Texas Law Review*, 78, 1, 105–47.

Solan, L. M. (2001). The written contract as safe harbor for dishonest conduct. *Chicago-Kent Law Review*, 77, 1, 87–120.

Solan, L. M. (2004). Pernicious ambiguity in contracts and statutes. *Chicago-Kent Law Review*, 79, 3, 859–88.

Solan, L. M. (2005). The new textualists' new text. *Loyola Law Review*, 38, 5, 2027–62.

Solan, L. M. (2009). Blame, praise, and the structure of legal rules. *Brooklyn Law Review*, 75, 2, 517–43.

Solan, L. M. (2010). *The Language of Statutes: Laws and their Interpretation*. Chicago, IL: University of Chicago Press.

Solan, L. M. and Tiersma, P. M. (2004). Author identification in American courts. *Applied Linguistics*, 25, 4, 448–65.

Solan, L. M. and Tiersma, P. (2005). *Speaking of Crime*. Chicago, IL: University of Chicago Press.

Stygall, G. (1994). *Trial Language: Differential Discourse Processing and Discursive Formation*. Amsterdam: John Benjamins.

Tiersma, P. (1987). The language of defamation. *Texas Law Review*, 66, 2, 303–50.

Tiersma, P. (1990). The language of perjury: literal truth, ambiguity, and the false statement requirement. *Southern California Law Review*, 63, 2, 373–431.

Tiersma, P. (1993). The judge as linguist. *Loyola Law Review*, 27, 1, 269–83.

Tiersma, P. (1999). *Legal Language*. Chicago, IL: University of Chicago Press.

Tiersma, P. (2001). The rocky road to legal reform: improving the language of jury instructions. *Brooklyn Law Review*, 66, 1081–118.

Tiersma, P. (2002). The language and law of product warnings. In J. Cotterill (ed.), *Language in the Legal Process* (pp. 54–71). Basingstoke: Palgrave Macmillan.

Tiersma, P. (2010). *Parchment Paper Pixels: Law and the Technologies of Communication*. Chicago, IL: University of Chicago Press.

Tiersma, P. and Solan, L. (2011). *The Oxford Handbook of Language and Law*. New York: Oxford University Press.

Wallace, W. D. (1986). The admissibility of expert testimony on the discourse analysis of recorded conversations. *University of Florida Law Review*, 38, 69–115.

32 Translation

CHRISTOPH GUTKNECHT

1 Introduction

Communication between different individuals and nations is not always easy, especially when more than one language is involved. The job of the translator and/or interpreter is to try to bridge the gap between two foreign languages. This can also include translation problems arising from historical developments within one language. In this chapter, translating and interpreting will be characterized as a communicative device (Section 2). After defining the modes of interpreting (Section 3), the principles influencing the transfer of messages from one language to another will be outlined (Section 4). In order to illustrate some of the lexical problems faced by the translator and/or interpreter, particular attention will be paid to the area known as "false friends" (Section 5). As will be emphasized, each act of translation is conditioned by many factors (with various functions) which govern the choice of a target-language rendition of a given source-language text (Section 6). In the final section, a brief survey of recent developments in machine translation will be presented (Section 7).

2 Translation: A Communicative Device

Translation is undoubtedly a communicative device; moreover, as John Rupert Firth (1956: 135) put it, "The fact is, translation is a necessity on economic and on general human grounds." Some researchers postulate an autonomous status for translation studies, arguing that these studies bring together work in a wide variety of fields, including literary study, anthropology, psychology, and linguistics. Others claim that the domain of translation studies is an important subbranch of applied linguistics. Proponents of both opinions would have to admit, however, that the field of translation studies has multidisciplinary dimensions and aspects.[1]

The term "translation" normally refers to written materials but is also an umbrella term used for all tasks where elements of a text of one language (the source language, SL) are molded into a text of another language (the target language, TL), whether the medium is written, spoken, or signed. There are specific professional contexts where a distinction is made between people who work with the spoken or signed language (interpreters), and those who work with the written language (translators). Although usually the two roles are seen as quite distinct, there are hybrid situations that blur this distinction. When, for instance, a court interpreter reads a legal document in one language while reciting it aloud in another s/he is said to be

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

sight-translating. On the other hand, prosecuting authorities and law enforcement agencies often call on translators to transcribe and translate foreign language conversations that were taped during investigations.

3 Modes of Interpreting: Consecutive and Simultaneous

There are two highly specialized modes of interpreting: *consecutive* and *simultaneous* interpreting.

One typically speaks of *consecutive interpreting* when the person requiring the interpreter participates in the communication directly. In such cases the interpreter waits for the person to finish speaking, or until the amount of information approaches the limit of the interpreter's retention capacity, and then the interpreter gives a translation. Interpreting skills include note-taking techniques, although the method and degree of reliance on note-taking as a memory aid varies from one interpreter to another. Consecutive interpreting is usually bidirectional, i.e., from language X to Y and vice versa; it is commonly used for informal meetings, tours, business negotiations, etc.

The mode of *simultaneous interpreting* is typically used when the person who requires an interpreter is not participating in the communication directly. At international conferences with bilingual or multilingual audiences simultaneous interpreting is an effective method for helping to overcome language barriers; it allows presentations and discussion to proceed at the same pace as an ordinary unilingual conference. Simultaneous interpreting is usually performed using technical equipment to relay the sound to those delegates who do not speak the floor language. The interpreters work in soundproof booths, while the delegates listen to the language of their choice via headsets connected to multichannel wireless receivers. In such cases the translation is usually unidirectional, i.e., from language X to language Y but not vice versa.

Strictly speaking, however, the term "simultaneous interpreting" is misleading in that the word "simultaneous" suggests that the interpreter is interpreting a message at the same time as hearing it. In fact, there is a delay between the moment the interpreter hears a number of SL expressions and the moment s/he renders them into the TL, because it takes time to understand the SL message and turn it into the TL. Meanwhile, the speaker goes on to the next utterance, so the interpreter must generate the TL version of the first utterance while processing the second, and so on. This delay is known as *décalage*, from the French word for "time lag."

In simultaneous interpreting, the time factor is generally more crucial than in consecutive interpreting or in translation performed in the written mode. The decisive factor in simultaneous interpreting is how early the simultaneous interpreter can actually start speaking. Wilss (1978: 346) says, quoting Mattern (1974: 28) in order to specify that moment:

> The optimal moment of interpretation will differ depending on the subjective and objective factors involved; the objective or speech-language-linked factors being those which originate from the SL text and from relations of equivalence existing between SL and TL, and the subjective factors being those which depend on the interpreter himself.

In view of the time pressure under which the simultaneous interpreter has to work, one crucial subjective factor is the interpreter 's memory. How long s/he is able to wait before s/he starts interpreting each sentence will depend upon the capacity to retain what was said by the SL speaker. S/he should be in a position to start interpreting as soon as possible to avoid being confronted with an information overload. But there are cases in which the simultaneous interpreter seems to be forced to wait for the conclusion of a long SL sentence before s/he can even start interpreting it. To evaluate this claim let us take a sentence from Wilss (1978: 348):

(1) a. *Namens meiner Fraktion darf ich den beiden Herren Berichterstattern für die Arbeit, die sie geleistet haben, sehr herzlich danken.*

It is true that, upon hearing the first three words, the German–English simultaneous interpreter can immediately start saying: "*On behalf of my political party*." On hearing the next two words s/he might add: "*I may*." Since the full verb follows the modal in English, the interpreter would have to wait until the very last German word of the complex sentence is uttered, i.e., the full verb *danken*. Such a late take-off or late continuation is "an extremely heavy stress on the short-term memory of the interpreter" (Wilss 1978: 347). Couldn't there be a solution providing a shortcut in this situation?

Mattern (1974: 3) suggests one such strategy, as reported by Wilss. If (1a) is uttered in an EU debate the experienced interpreter will know that

> the German segment "*Namens meiner Fraktion darf ich* (*danken*)" is a standard phrase which is frequently used as an opening gambit in a follow-up speech statement . . . Once the simultaneous interpreter has heard "*Namens meiner Fraktion darf ich*," he can legitimately infer from previous experience that some form of saying "thank you" can be expected. (Wilss 1978: 348)

This is why, having heard only the first five words, the interpreter can start or continue his TL rendition.

This form of "intelligent textual prediction," referred to by Wilss (1978: 348ff.) as "syntactic anticipation," has yet another advantage. It saves the interpreter from mistakenly rendering the German *dürfen* by the English *may*, which would be appropriate in many other cases. Instead *would like to* has to be employed, as the following complete rendition of (1a) shows:

(1) b. *On behalf of my political group, I would like to thank the two spokesmen very cordially for their work*.

Even though *may* is in many cases the adequate rendition of *dürfen*, in the present case this principle has to be revised due to the modal's being part of the syntactic construction "*darf . . . danken*" which has to be considered as a whole. This construction may therefore be called a "revision factor" – the standard equivalence (English *may* = German *dürfen*) does not apply here but has to be revised. This shows that it is by no means invariably sufficient to go by a modal alone in order to render it; rather, there are different translation units which may be relevant to different SL texts.

Syntactic anticipation is but one of the interpreter's skills. Both interpreters and translators must be skilled in such generalized professional techniques, and they must also be intimately familiar with the material under discussion in a given text. This holds true for all types of interpretation and translation but is especially important in the areas of technology, medicine, and the law, where terminological accuracy is of paramount importance. Because of the rapid development of science and technology having resulted in a significant increase in the amount of knowledge being transferred across languages and cultures, it is imperative for any translator to have access to multilingual terminology databases.

For court interpreters, who have to deal with lawyers, court personnel, and the public, it is imperative that they have an understanding of the terminology and procedures used in court as well as an extensive vocabulary ranging from formal legal language to colloquialisms and slang expressions.

In the last few years there has been an increasing awareness of the importance of using trained professionals rather than well-intentioned amateurs for community interpreting. Many hospitals, courts, and other institutions in the USA (but fewer in Europe) now have staff positions for interpreters and translators to aid immigrants in communicating with and in gaining equal access to legal, health, and social services. It is especially for these positions that interpreters are required who are able to act as paranative speakers and biculturalists and are aware of the fact that "no language can impartially transmit information independently of particular forms of culture and knowledge . . . [and that] English therefore inevitably carries the biases and presuppositions of certain cultural traditions" (Hyland 1997: 20).

4 Translation Principles

Much valuable work has been carried out by translatologists on the methodical scrutiny of translation, establishing interesting but often contradictory translation principles. Savory's (1968: 54) collection includes the following:

- a translation must give the words of the original;
- a translation must give the ideas of the original;
- a translation should read like an original work;
- a translation should read like a translation;
- a translation may add to or omit from the original;
- a translation may never add to or omit from the original.

The idea underlying these statements is to postulate what is "right," but contradictory statements such as these can obviously not all be right at the same time. However, each of these postulates can be valid in its own right. To take an extreme interpretation of the first pair of principles as an example: The demand that "a translation must give the words of the original" preserving the successive units of the source text and arranging them in order of occurrence irrespective of the "normal" grammatical sequence of units in the TL (i.e., an *interlinear translation*) is justifiable *if* the aim is to carry out comparative linguistic research:

On the word level, Dutch *daarmee*, for instance, corresponds to English *therewith*, cf.

(2) a. *Daarmee hebben we het gedaan, niet met de hamer.*
b. *There with have we it done, not with the hammer.*

The English sentence gives "the words of the original." *If*, however, the main purpose of a translation is to describe to the hearer a certain state of affairs as closely as possible, then "a translation must give the ideas of the original." Along these lines, Hannay (1989: 224ff) points out that the standard translation of Dutch *daarmee* "is not the archaic *therewith* but *with it/with that/with them*, depending on the nature of the referent . . .:

(2) c. *'That's what we did it with, not the hammer.'* "

Note the two "ifs" used above: The statements just made are no longer as absolute and unconditional as those quoted by Savory; but rather, they are made relative to different target factors specifying the purpose of the translation. In this way the age-old question whether a translation should be literal ("word for word") or free ("sense for sense") is no longer a matter of controversy – it turns out to be not so much a question of arguments to be adduced for deciding which of the two principles is better or right per se; rather, opting for one or the other of the two principles – and, indeed, for any translation principle – is a matter of clear-cut requirements relative to a given purpose or target resulting from a specific commission. Or, to put it differently, the arguments in favor of each principle result from a set of factors that were previously defined. Taking these factors to constitute counterarguments against (an)other principle(s) is a futile endeavor since it is of no relevance to practical translation work.

Hervey, Higgins, and Loughridge (1995: 43) referred to an extraordinary example of a translation where the sound of the source text was chosen to be the decisive factor "allowing the sense to remain at best a vague and suggested impression." Here is part of one of Catullus' (Latin) poems (3a), followed by Celia and Louis Zukovsky's "phonemic translation" (3b). They are attempting to replicate in their rendition the sound sequence of the source text:

(3) a. Ille mi par esse deo videtur,
ille, si fas est, superare divos,
qui sedens adversus, identidem te spectat et audit
dulce ridentem, misero quod omnis eripit sensus mihi; . . . .

b. He'll hie me, par *is* he? The God divide her,
he'll hie, see fastest, superior deity,
quiz – sitting adverse identity – mate, inspect it and audit –
you'll care ridden then, misery holds omens,
air rip the senses from me; ....

What becomes obvious here is that this "translation" sets out to imitate as closely as possible the actual sound sequences of the original, while the content is only vaguely incorporated in the English rendition.

Be it such a phonemic translation, or a word-for-word translation, or a free translation – preferring one principle to another one is a matter of relevance to the target group as viewed by the client. Gutt (1991: 121) says that "the different 'translation principles' do reflect differences in what different readers consider to be relevant . . . Thus the contradictions can be resolved when each principle is not stated in absolute terms, but qualified by the condition: 'when required for consistency with the principle of relevance.'"

As we have just seen in the above example, the most important factor is not always required to be the original meaning of the text, but can be, for example, the original sound. This leads us to another important factor that involves sound – words that are identical or at least very similar in spelling and/or sound in two or more languages. I am referring, of course, to "false friends."

5 False Friends

You come across false friends more often than you would like to – not only in real life, but also in linguistics, especially when you happen to be doing a translation.

When someone refers to the so-called "translator 's false friends," s/he means the English adaptation of *faux amis du traducteur*, a French expression that has been used since 1928, when Maxime Kœssler and Jules Derocquigny published a book in Paris with the title *Les Faux Amis ou les trahisons du vocabulaire anglais* ("False friends or the treacherous pitfalls of the English vocabulary").

The fact that "false friends" sound alike often leads to the incorrect assumption that they have the same meaning; however, that is sometimes only partially the case, and often not at all.

We can safely say that these false friends are a serious linguistic problem which belongs to the field of *interference* (sometimes also called *negative transfer*). *Interference* is the phenomenon that we experience when linguistic structures that we have already learnt interfere with our learning new structures. Interference between two languages exists in all areas – for example, in pronunciation and spelling. Incidentally, interference exists not only between two languages, but also within one language. In semantics, one therefore refers to *intralingual* and *interlingual false friends*. Since a word may change its meaning in the course of time, this problem cannot be viewed only in the light of the current (i.e., *synchronic*) situation. Because the historical (i.e., *diachronic*) development must also be taken into consideration, there are altogether four types of false friends.

At this point it might be interesting to look at some illustrative examples of how the meanings of words can be confused because of misleading similarities in two languages. In the examples, the language pairs German–Italian, English–Italian, and English–French will be used.[2]

5.1 Synchronic interlingual false friends

Ronnie Ferguson, author of *Italian False Friends* (1994: ix), rightly emphasizes that "[a]ccurate translation . . . as well as the proper appreciation of advanced Italian texts, hinge on the confident handling of key words"; among other examples he mentions key words such as *attuale* (*present/topical*, never *actual*), and *eventuale* (not *eventual* but *possible*), which – like their German "true friends" (*aktuell* and *eventuell*) – are false friends of the English words resembling them in form.

Similarly, anyone who would translate the German *luxuriös* and the English *luxurious* with the Italian expression *lussurioso* would be committing a big *faux pas*. The correct translation would be *lussuoso* and not the Italian word *lussurioso*, which has the same meaning as the German word *lasziv* and the English *lascivious*.

5.2 Diachronic intralingual false friends

We can encounter *diachronic intralingual false friends* if we translate linguistic elements from one historical period into another period, when the process of shift in meaning has to be taken into account. A spectacular case in point is the word *nice*: In Old French, which gave the word to English in the thirteenth century, it meant "simple," "silly," and in turn was based on Latin *nescius*, which meant "ignorant." In the fourteenth century, *nice* in English acquired the meaning of "wanton," "loose-mannered," even "lascivious." This sense occurs, for instance, in line 1285 of Geoffrey Chaucer 's "The Romaunt of the Rose" (1366): "Nyce she was, but she mente Noone harme ne slight in hir entente, But oonely lust & jolyte." So translating Chaucer 's *nice* with the modern English *nice* (meaning "friendly, agreeable, pleasing") would be incorrect.

5.3 Diachronic interlingual false friends

Since language changes constantly, the meaning of expressions can broaden as well as narrow down, and can denote something "better" as well as something "worse." For this reason, words in two languages that were originally true friends can develop into false friends (and vice versa). Carlo Milan highlighted *diachronic interlingual false friends* in an essay in a 1989 volume of the journal *Sprachwissenschaft*, where he compared the German words *Artist* and *realisieren* with their Italian counterparts. He pointed out that the German word *Artist* was derived from the French expression *artiste*, meaning "artist" in the general sense of "somebody performing an art." However, the meaning of this expression was gradually narrowed down in German to *Artist* in the sense of "acrobat," and thus became a false friend because the Italian word *artista* has preserved its original meaning (and can even be modified, such as in *artista di circo* or *artista di varie*tà); the correct modern German equivalent of the Italian word *artista* (denoting "somebody performing an art") is *Künstler*. We can see then that a gradual intralingual change in meaning leads to the creation of interlingual false friends.

As indicated above, words that are false friends at a certain point in time can later become true friends. One factor that plays a decisive role in this change is the increasing tendency to internationalize certain words which sound the same in two or more languages although they originally had (at least partially) different meanings. An interesting example is the German *realisieren* and the Italian *realizzare*, which were originally both used exclusively in the sense of "to realize profits, projects, hopes or dreams" or "to make." Both words were greatly influenced by the English *realize* in that their meanings today also include "to comprehend."

In their book *Faux Amis and Key Words*: *A Dictionary-Guide to French Language, Culture and Society through Lookalikes and Confusables* (1985), Philip Thody and Howard Evans pointed to a similar development. They commented on the English equivalents of the French verb *réaliser*, stressing that the verb originally meant " 'to achieve (one's ambition),' 'to realize (one's assets),' but not – at least for the purists – 'to realize (become aware),' which is *se rendre compte que* or *de*, though most French people do, in conversation, use *réaliser* in the latter sense" (1985: 78). We can therefore conclude that the English verb *realize* is in the process of becoming truly international.

5.4 Synchronic intralingual false friends

Further problems for the translator are caused by *synchronic intralingual false friends*, even by native speakers. For instance, in German one has to distinguish between *fremdsprachiger Unterricht*,

i.e., "teaching in a foreign (or target) language" and *fremdsprachlicher Unterricht*, i.e., "foreign-language teaching."

Many English words that appear to mean the same can also lead to confusion and make life very difficult for translators and interpreters. It would be extremely dangerous, for example, to assume that *inflammable* is the opposite of *flammable*; in fact, both words mean the same. The 1992 edition of the *BBC English Dictionary* defined the two words as follows: "An *inflammable* material or chemical 'burns easily.' – Something that is *flammable* 'catches fire easily.' "

Nevertheless, even in England, *inflammable* is incorrectly used in the sense of "nonflammable" because many people believe that the *in-* at the beginning of the word gives it a negative meaning, similar to the *in-* at the beginning of *incomplete* and *indirect*. In an attempt to avoid any grave errors, the British Standards Institution issued the following warning in 1959: "It is the Institution's policy to encourage the use of the terms *flammable* and *non-flammable* rather than *inflammable* and *non-inflammable*."

The fact that even in England often the wrong meaning "nonflammable" was ascribed to the word *inflammable* shows that the word *inflammable* is both an intralingual and an interlingual false friend.

We have seen that there is a host of factors affecting the act of translation and that the translator should take them all into consideration when translating – in other words, s/he should "translate by factors."

6 Translating by Factors

Each act of translation is conditioned by a huge variety of factors – factors that can and must be identified for the act of translation to be taught, learned, and practiced.[3]

A worthwhile endeavor of translation theory would be to do research into the ways and means of creating optimum TL renditions of (different kinds of) SL texts in the light of different factors which have to be taken into account.

As was pointed out in Section 3, rendering the German modal verb *dürfen* by English *may* is adequate in many cases but inappropriate in the context where sentences (1a) and (1b) are uttered. In Section 5, another "revision factor" was mentioned which comes into play when the translator wrongly believes in false friends, words which look or sound similar but do not have an identical meaning.

As was shown, translating by factors does not mean taking prescribed factors into account as such but bearing in mind the specific roles or functions these factors fulfill. Gutknecht and Rölle (1996: 5ff.) list a number of basic functions of translation factors. Besides revision factors as exemplified above, every translator/interpreter is, for instance, faced with "blocking factors" on which the revision factors are based. These blocking factors make a specific TL rendition impossible. An example will make this point clear.

An indefinite number of English combinations of the form *adjective plus noun* can safely be translated into German:

red rose	*rote Rose*
interesting film	*interessanter Film*
beautiful house	*hübsches Haus*

In English, the expression *simultaneous interpreter* may also be added to the list, but in German the structurally corresponding construction *simultaner Dolmetscher* is not possible. The reason is that it would suggest that the interpreter himself is simultaneous. In German the rule is that for an adjective to premodify a noun it must denote a characteristic of the referent of that noun; otherwise the formation will be ungrammatical. So in the case of the SL expression *simultaneous*

interpreter a semantic factor (viz., the information that *simultaneous* is no characteristic feature of *interpreter*) will act or function as a blocking factor to the TL rendition **simultaner Dolmetscher*: the correct German version would be *Simultandolmetscher*.

In addition to revision factors and blocking factors, there are, among others, "invariance factors," which make an SL feature reappear in the TL rendition (e.g., the English expression *This book is a must* can be rendered into German by *Dieses Buch ist ein Muss*). "Change factors" make an SL feature disappear or a new or additional feature appear in the TL rendition: In German *ein Muss* cannot be pluralized the way *a must* can, such as in *three important musts*. Change factors often become effective in everyday language and in specialized communication whenever concepts from different languages differ considerably in their characteristics, or when a concept exists in only one language. Lynne Bowker (1994: 184) outlined "five strategies for handling such 'untranslatable concepts': use of footnotes, use of the closest corresponding TL equivalent, paraphrasing the SL term, use of loan words and loan translations, and creation of neologisms."

"Target factors" relate to the target or purpose of the translation as determined by the client, for instance, carrying out comparative linguistic research or describing to the hearer a certain state of affairs as closely as possible (see Section 4). It is the client who makes basic choices in each act of commissioning and who prescribes how to translate. For instance, the client is the one to decide how faithful or how free the translation is to be, whether the target is an interlinear translation, a phonemic translation (both exemplified above), or another kind of rendition.

7 Machine Translation and Computer-Assisted Translation

Because factors are objectifiable, the factor approach is an ideal tool for machine translation. Machine-aided human translation (MAHT) is to be distinguished from fully automatic machine translation (FAMT). MAHT, also known as computer-assisted translation (CAT), involves some interaction between the translator and the computer. In contrast, FAMT, better known as machine translation (MT), is characterized by the absence of any human intervention during the translation process. Judith Klavans (in O'Grady, Dobrovolsky, Aronoff, and Rees-Miller 1997: 656) rightly emphasizes that

> the purpose of a machine translation system is the same as that of any translation system: taking text written or spoken in one language and writing or speaking it in another . . . Translation poses challenging problems both for the human translator and for the machine attempting to do what the human does.

Tests conducted at the offices of many international organizations, for instance at the Pan American Health Organization, the WHO Regional Office for the Americas, have demonstrated that in its present stage of development, fully automatic translation technology is not considered cost effective because its resulting output needs extensive revision work (postediting). The machine-aided human translation approach, on the other hand, seems to be more suited to the needs of many organizations which have to handle the translation of documents. Computer-assisted translation systems are based on "translation memory." With such systems (that are sometimes combined with terminology databases), translators have immediate access to previous translations of portions of the text, which they can then accept, reject, or modify. By constantly archiving their final choice, translators will soon have access to an enriched "memory" of ready-made solutions for a wealth of translation problems. Other recent developments in computer technology also help the translators to perform their job. There is, for instance, a new and very effective productivity tool available for PC-based translators: automatic dictation software. At the present state of speech-recognition technology, however, to use dictation effectively the translator must master a new foreign language: "paused" speech that–the–computer–can–understand.

Regardless of the degree of usefulness of machine translation, there seems to be unanimous agreement that translators cannot be replaced, either now or in the foreseeable future. Only the human translator as an expert will be able to fully survey all the factors relevant to felicitous translation processes.

RELEVANT JOURNALS

For further reading on human and machine translation, consult the following translation periodicals: *Babel* (Amsterdam), *Interpreting* (Amsterdam), *Language International* (Amsterdam), *Lebende Sprachen* (Berlin), *Machine Translation* (Dordrecht/The Netherlands and Norwell, MA), *Meta* (Montreal), *Multilingua* (Berlin and New York), *Perspectives* (Copenhagen), *Target* (Amsterdam), and *The Translator* (Manchester).

NOTES

1 Two useful collections of readings on general aspects of translating and interpreting are Owens (1996) and Sofer (1997); for a comprehensive overview of translation studies as an academic discipline, see Baker (1997).

2 From the mass of literature dealing with false friends, I select the following for additional mention:
   - The *Cambridge International Dictionary of English* (1995) is unique in that it uses special symbols to warn its consulters about false friends and contains lists of false friends for 16 languages (Czech, Danish, Dutch, French, German, Greek, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Russian, Spanish, Swedish, and Thai).
   - For the English-speaking translator of German, Fred Bridgham (1996) is especially interesting since his explanations are illustrated with examples from literature, the press and everyday life.
   - Henri van Hoof (1989) offers most valuable hints for the English translator of French.

3 For an attempt to verify this statement, see Gutknecht and Rölle (1996: chs 2–7) and Gutknecht and Rölle (2001).

REFERENCES

Baker, M. (ed.) (1997). *Routledge Encyclopedia of Translation Studies*. London and New York: Routledge.

BBC English Dictionary (1992). (Editor-in-chief, J. Sinclair). London: HarperCollins.

Bowker, L. (1994). Applied terminology: a state-of-the-art report. *Terminology*, 1, 1, 181–92.

Bridgham, F. (1996). *The Friendly German–English Dictionary. A Guide to German Language, Culture and Society through Faux Amis, Literary Illustration and Other Diversions*. London: Libris.

Cambridge International Dictionary of English (1995). (Editor-in-chief, P. Procter). Cambridge: Cambridge University Press.

Ferguson, R. (1994). *Italian False Friends*. Toronto: University of Toronto Press.

Firth, J. R. (1956). Linguistic analysis and translation. In M. Halle, H. G. Lunt, H. McLean, and C. H. van Schooneveld (eds), *For Roman Jakobson: Essays on the Occasion of his Sixtieth Birthday 11 October 1956* (pp. 133–9). The Hague: Mouton.

Gutknecht, C. and Rölle, L. J. (1996). *Translating by Factors*. Albany, NY: State University of New York Press.

Gutknecht, C. and Rölle, L. J. (2001). Translation factors. In W. Thiele, A. Neubert, and C. Todenhagen (eds), *Text – Varieties – Translation* (pp. 25–42). Tübingen: Stauffenburg.

Gutt, E.-A. (1991). *Translation and Relevance: Cognition and Context*. Oxford: Blackwell.

Hannay, M. (1989). Translating structures: the role of contrastive syntax in translation dictionaries. In J. L. Mackenzie and R. Todd (eds), *In Other Words: Transcultural Studies in Philology, Translation and Lexicology Presented to Hans Heinrich Meier on the*

Occasion of his Sixty-Fifth Birthday (pp. 211–34). Dordrecht: Foris.

Hervey, S., Higgins, I., and Loughridge, M. (1995). *Thinking German Translation: A Course in Translation Method: German to English*. London: Routledge.

Hoof, H. van. (1989). *Traduire: L'anglais: théorie et pratique*, 2nd edn. Paris: Editions Duculot.

Hyland, K. (1997). Scientific claims and community values: articulating an academic culture. *Language and Communication*, 17, 19–31.

Mattern, N. (1974). Anticipation in German–English simultaneous interpreting. MA thesis, Saarbrucken.

Milan, C. (1989). Falsche Freunde. Ein besonderes Problem der kontrastiven Lexikologie (Deutsch–Italienisch). *Sprachwissenschaft*, 14, 384–404.

O'Grady, W., Dobrovolsky, M., Aronoff, M. , and Rees-Miller, J. (eds) (1997). *Contemporary Linguistics*, 3rd edn. New York: St. Martin's Press.

Owens, R. (ed.) (1996). *The Translator's Handbook*, 3rd edn. London: Aslib.

Savory, T. H. (1968). *The Art of Translation*. London: Cape.

Sofer, M. (ed.) (1997). *The Translator's Handbook*. Rockville, MD: Schreiber.

Thody, P. and Evans, H. (1985). *Faux Amis and Key Words: A Dictionary Guide to French Language, Culture and Society through Lookalikes and Confusables*. London: Athlone Press.

Wilss, W. (1978). Syntactic anticipation in German–English simultaneous interpreting. In D. Gerver and H. W. Sinaiko (eds), *Language Interpretation and Communication* (pp. 343–52). NATO Conference Series 3: Human Factors 6. New York and London: Plenum.

33 Language Planning and Policy

KATHRYN D. STEMPER AND KENDALL A. KING

1 Introduction

The field of language planning and policy (LPP) is concerned with the policies – both explicit and implicit – that influence what languages are spoken when, how, and by whom, as well as the values and rights associated with those languages (or language varieties). Bernard Spolsky (2012) offers a three-tiered definition; in his view, language policy encompasses (1) language practices, (2) language beliefs or ideologies, and (3) efforts to modify or influence those within a particular community. Spolsky maintains that *language policy* is the broader term, encompassing all three aspects (practices, beliefs, and efforts), with *language planning*, in turn, focusing on the third component (efforts to modify). This is also the position taken in this overview.

LPP is an exciting area of study because language policies – as well as how they are implemented and resisted – are often sources of political debate, both reflecting and contributing to social, economic, and cultural tensions across the broader society. For instance, while the USA has no official national language (Spolsky 2004), state and federal language policies prescribe what languages are used in schools for what purposes as well as whether and how students might have opportunities both to use *and* maintain their native (home) languages. The changing nature of these policies largely reflects national sentiment toward immigration and diversity. As an example, the 1968 Bilingual Education Act (BEA) (signed into law as Title VII of the Elementary and Secondary Education Act [ESEA]) was the first federal education policy that supported instruction through the medium of students' native languages, and was reflective of the more liberal and progressive politics of then president Lyndon B. Johnson's administration (San Miguel 2004). This legislation has since been diminished through federal policy as changes to the ESEA have placed greater focus on *English* acquisition rather than bilingual education. Additionally, as concerns over immigration and shifting demographics increased in the late 1990s, policies that further limited opportunities for bilingual education were implemented at the state level. Proposition 227 in California (1998) and Proposition 203 in Arizona (2001) are examples of two initiatives that put strict limits on the amount of bilingual support that non-native English speaking students receive. Despite abundant research evidence supporting the efficacy of bilingual education (Wright 2005), these measures effectively banned bilingual education programs in those states.

Another recent example of LPP in action is evident in the ongoing debates concerning the medium of instruction (MOI) in Hong Kong, where English had long been used as the language of instruction in secondary schools, despite the fact that many students lacked adequate

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.
© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

competency to do advanced academic work in English. In 1997, Hong Kong's Department of Education outlined a new government education policy that called for a change in MOI, from English to most students' mother tongue, (Cantonese) Chinese, in secondary schools unless schools could prove that teachers and students were able to adequately teach and learn through English. This policy was based on the recognized benefits of mother-tongue education as well as research reports suggesting that English MOI negatively influenced students' learning. At the beginning of the 1998–9 school year, 70 percent of secondary schools had changed English to Chinese MOI (Zeng 2007). This change was met with great resistance, with many parents fearful that their children would no longer be competitive in the labor market or higher education without English fluency, which they believed would be best achieved through English-medium schooling. As Zeng states, this "show[s] that language in education policy is not only a school, but also a social issue" (p. 43).

These two examples are both cases of so-called *top-down* language policy in that they are written and implemented through authoritative bodies that are part of federal or national government (e.g., departments or ministries of education). However, language policies can also be *bottom-up*, in other words, constructed by communities to address salient language issues. For instance, in the 1980s, Indigenous groups in Ecuador developed language policies to support Quichua and other Native languages, and advocated powerfully and successfully for their adoption (King and Haboud 2002). LPP scholars focus on understanding the development of both *top-down* and *bottom-up* language policies (Hornberger 1997), but also their implementation and impact at local, regional, and national levels. While the focus of LPP is on how language policies are formed, implemented, and sometimes resisted, LPP scholars note that nonlinguistic issues often play a critical role. For instance, as suggested above, attitudes about immigration and national identity have been crucial in shaping US language and education policy. Additionally, although explicit language policies certainly exist, many are implicit and reflect, as Spolsky (2012) states, the beliefs and practices of a language community. A prime example: while the United States does not have a policy declaring English the country's official language, there is, however, no question that English is the language of government, power, and social mobility. Below, we provide an overview of how LPP has developed as an academic field over time, of current research and theories in the field, and of current areas of investigation.

2 Central Concepts and Questions

2.1 Definitions: language policy vs. language planning

At its inception in the late 1950s, the field was known as *language planning*. Early work was conducted in multilingual European countries such as Sweden and Norway, but often concentrated on the language policies of newly independent nations such as Indonesia and India (Fishman et al. 1968). As these multilingual, former colonial states set up their own governments, complicated and important questions emerged. These included: "What should be the official and national languages?," "What languages should be taught as subjects and which used as media of instruction?," and "What varieties should be used and how should written standards be developed?"

Some of this work focused on corpus planning (which concerns the structure of language) and other work on status planning (which concerns the functions of language). As linguists, sociologists and anthropologists increasingly engaged with these issues, Einar Haugen (1959: 8) famously (although in retrospect, somewhat narrowly) defined language planning as "the activity of preparing a normative orthography, grammar, and dictionary for the guidance of writers and speakers in a non-homogeneous speech community." Haugen's early work, which concentrated on corpus planning, focused on the languages of Norway and explored the transformation of Danish language varieties into what is now considered standard Norwegian.

In more recent decades, the term *language planning* has often been used – sometimes interchangeably – with *language policy*. Although there has been some debate about the relationship between "language policy" and "language planning" and myriad attempts to cleanly differentiate them, language policy and language planning are, in fact, "inextricably related" (Hornberger 2006: 25). There is no widely agreed upon, cut and dry distinction between *planning* and *policy*. For instance, some argue that planning subsumes policy (Fettes 1997; Kaplan and Baldauf 1997), while others maintain that policy subsumes planning (Ricento 2000; Schiffman 1996). By some accounts, the act of language planning leads to language policy (Spolsky 2012: 3). Although the terms tend to reference different activities (Johnson 2013: 3), the field is now commonly referred to as language planning and policy (LPP) and has been adopted as the default term encompassing all of these activities. Spolsky's three-tiered definition, described at the start of this chapter, at least at the present moment, is the most commonly used.

2.2 *Areas of language policy and planning: corpus, status, and acquisition*

As noted above, there is a longstanding distinction in the field between corpus and status planning. Presently, researchers tend to distinguish among three areas of LPP: corpus, status, and acquisition. See Table 33.1 for examples of each.

Kloss (1969) introduced the distinction between corpus planning and status planning in his report about "the possibilities for research work in the areas of institutional and of socio-cultural bilingualism" in Canada (p. 1). Similar to Haugen's definition of language planning, Kloss stated that

> planning with regard to languages is usually understood to mean that some agency, person, or persons are trying to change the shape or the corpus of a language by proposing or prescribing the introduction of new technical terms, changes in spelling, or the adoption of a new script. (p. 81)

Table 33.1 Areas of language policy: examples of and suggestions of further reading about status, corpus, and acquisition planning.

	Focus on	Example	Read more here
Status	Uses of language	In 2010, through voter referendum, Oklahoma declared English its official language, thus becoming the 31st state in the USA to declare English as its official language.	Hornberger and King (1996) Hult (2004) Brown (2013)
Corpus	Language	In 2006 the Council for German Orthography removed controversial changes of spelling and grammar from the German Orthography Reform of 1996.	Haugen (1959) Haugen (1983) Johnson (2005) Kloss (1969)
Acquisition	Users of language	Education policies in the USA, at both the state and federal level, impact bilingual education opportunities in schools, thus determining what languages students can learn and use in schools.	Cooper (1989) Menken (2008)

Per this definition, he suggested that when language planning is centered on modifying the language itself it is corpus planning. Status planning concerns questions such as "Which languages should be used when and for what purposes?" Kloss acknowledged the importance of status planning as well, noting that "there exists, however, another dimension of planning where one busies oneself not with the structure and form of language but with its standing alongside other languages or vis-a-vis a national government" (p. 81). Language planning efforts dedicated to status planning might attempt to raise or lower the status of a language. For instance, the status of Sami was raised when recognized as an official language of Sweden in 1998/9 per the *National Minorities in Sweden* bill (Hult 2004). Cooper (1989) introduced *acquisition planning* as planning which "is directed toward increasing the number of users – speakers, writers, listeners, or readers" (p. 33). Schools or other institutions where opportunities for language learning are created (or limited) are often the setting of acquisition planning.

While the descriptions of status, acquisition, and corpus planning can make them seem like very distinct activities, Fishman (1979) illustrated how these areas of LPP intertwine, with corpus planning often reflecting and guided by the status planning environment. In Fishman's words, "status planning without concomitant corpus planning runs into a blind alley. Conversely, corpus planning without status planning is a linguistic game, a technical exercise without social consequence" (p. 12). For instance, in the case of Indigenous language revitalization, promoting use of Indigenous languages in education (status and acquisition planning) often leads to debates about the structure of the language and a perceived need for standardization (corpus planning). Writing about Hawaiian language revitalization, Wong (1999) describes the interplay and occasional tension between acts of status planning (allowing for immersion education in schools) and corpus planning (modernization of the Hawaiian language to meet the needs of curriculum).

2.3 Historical phases of LPP

Since the pioneering early work of Einar Haugen, Charles Ferguson, Joan Rubin, and others, the field has undergone a number of important shifts. Ricento (2000) defined three historical phases of LPP, each lasting roughly two decades. These phases have been shaped by large-scale world events, such as decolonization in the 1960s and globalization in recent decades, but also reflect trends in the social sciences, including evolving notions of appropriate research questions and what methods are best to answer these questions. Each of these LPP phases is briefly outlined below.

2.3.1 Phase 1 As suggested above, as the field of LPP emerged in the late 1950s and early 1960s, it often focused on solving "language problems," typically in newly independent nations. Indeed, a foundational text of the time was called *Language Problems of Developing Nations* (Fishman, Ferguson, and Das Gupta 1968). Some early researchers focused on corpus planning, which was considered to offer "solutions" to the so-called "language problems." Ricento refers to this early approach to solving language problems through language planning as *classic language planning* (2000: 206).

2.3.2 Phase 2 Beginning in the late 1970s, this "classic" approach attracted growing criticism as it became clear that the language planning "solutions" were not always effective, and indeed, often supported established power structures. Researchers increasingly began to focus on the social, economic, and political effects of language planning. Wolfson and Manes (1985), for instance, in their edited volume *Language of Inequality*, drew attention to the ways in which language both reflects and influences "social, economic, and political inequality" (p. ix). Critical theories within sociolinguistics entered the field through the work of Cooper (1989) and Tollefson (1991) as research began to highlight the negative impact of language planning, for instance, the

inequalities produced as a result. James Tollefson tirelessly promoted critical theory in LPP (1991, 2000, 2002, 2006, 2013), and (re)defined language policy as "an institutionalization of language as a basis for distinctions among social groups" (1991: 16).

2.3.3 Phase 3 The line between the second and third (current) phase is blurry; however, Ricento (2000) suggests that it started in the mid 1980s as new language policy topics emerged and research became increasingly *overtly* political. Cibulka (as cited in Ricento 2000: 197) observes that "the borderline between policy *research* and policy *argument* is razor thin." An important work in this vein was Phillipson's book, *Linguistic Imperialism* (1992), in which he argues that Western countries and institutions, such as the World Bank and British Council, have used English (and the promotion of English) as a means to continue to dominate former colonial nations. Critical theory that focuses on power relationships in LPP continues to be a focus for researchers. Additionally, issues such as language loss (particularly loss of Indigenous languages) have also gained attention and urgency (King 2001; King et al. 2008).

Moreover, as evident in the next section, in contrast to early scholarship in the field, which tended to focus on national-level overviews of LPP, more recent research has give more attention to examining site-specific interactional processes and discourse patterns, and their attendant ideologies. Sites of active investigation include not just nations or states, but LPP in the home (see Section 3.1), in schools, within the media, and in workplaces. Current research in LPP also increasingly focuses on how LPP are implemented and negotiated by targets potentially disrupting the original intention of top-down policies. Current research also takes a critical stance on central constructs (e.g., "diglossia," "native speaker," "language") with a keen eye to how LPP produces, exacerbates, or occasionally ameliorates inequality (Romaine 2006; Makoni and Pennycook 2005). In this work, a focus on *language ideologies* suggests connections between common language ideologies and attitudes, and effects of language policies on marginalized groups (King and De Fina 2010).

2.4 LPP goals and frameworks

While LPP is a relatively young field of study, it is ripe with frameworks. As the field moved away from "solving" "language problems," Cooper (1989: 98) posed the following question as central to LPP: "What *actors* attempt to *influence* what *behaviors* of which *people* for what *ends* under what *conditions* by what means through what *decision-making process* with what *effect*?" The following frameworks provide a means of analyzing LPP goals and processes.

2.4.1 Integrated framework for LPP Hornberger's *Integrated Framework for Language Policy and Planning Goals* (1994, 2006) draws from and integrates early definitions of LPP goals, aims, and objectives, including those put forth by Cooper (1989), Haugen (1983), Ferguson (1968), Hornberger (1994), Kloss (1968), Nahir (1984), Neustupny (1974), Rabin (1971), and Stewart (1968) (see Table 33.2). The framework brings together a range of LPP goals (e.g., officialization, maintenance) and juxtaposes the three types of LPP (status planning, acquisition planning, and corpus planning) across two different approaches: the policy planning approach (focusing on form) and the cultivation planning approach (focusing on function). Hornberger (2006) explains, "the policy planning approach attends to matters of society and nation at the macroscopic level and concerns standard language" and is usually associated with standard language use. In contrast, "the cultivation-planning approach deals with matters relating to language/literacy at the microscopic level and is concerned with literary learning" and is usually associated with ways of speaking and writing (p. 28).

Hornberger and King (2006) utilized this framework to demonstrate aspects of corpus, status, and acquisition planning based on King's ethnographic study of a community-based,

Table 33.2 Language policy and planning goals: an integrative framework. Reprinted from *An Introduction to Language Policy: Theory and Method Book Title* (p. 29), by T. Ricento (Ed.), 2006, Malden, MA: Blackwell. Reprinted with the permission of Wiley.

Types	Policy planning approach (on form)	Cultivation planning approach (on function)
Status planning (about uses of language)	Officialization	Revival
	Nationalization	Maintenance
	Standardization of status	Spread
	Proscription	Interlingual communication
		– International
		– Intranational
Acquisition planning (about users of language)	Group	Reacquisition
	Education/school	Maintenance
	Literary	Shift
	Religious	Foreign language/second language/literacy
	Mass media	
	Work	
	Selection	Implementation
	Language's formal role in society	Language's functional role in society
	Extra-linguistic aims	*Extra-linguistic aims*
Corpus planning (about language)	Standardization of corpus	Modernization (new functions)
	Standardization of auxiliary code	– Lexical
		– Stylistic
	Graphization	Renovation (new forms, old functions)
		– Purification
		– Reform
		– Stylistic simplification
		– Terminology unification
	Codification	Elaboration
	Language's form	Language's functions
	Linguistic aims	*Semi-linguistic aims*

Indigenous-led effort to incorporate Quichua as a second language in an Indigenous school of the Ecuadorian highlands.

- *Status planning.* When space was made in the (previously all Spanish-language) school for Quichua, "the functional allocation" of the language (p. 438) was altered. This effectively raised the status of Quichua, thus contributing to Quichua revitalization as cultivation planning.
- *Acquisition planning.* One of the primary aims of the school was to promote usage of Quichua through instruction (policy planning approach), which would also impact goals of Quichua revitalization and reacquisition in the community (cultivation planning approach).
- *Corpus planning.* There are multiple regional varieties of Quichua in the country, while Unified Quichua is used for educational texts and literature. The use of Unified Quichua for language instruction created, or at least accentuated, "generational and educational differences within

the community" (p. 433). However, as Hornberger and King state, "the impact the initiatives have in promoting standardization and unification of the varieties [policy planning approach] must be seen as a positive factor in language revitalization" (p. 438).

2.4.2 Critical framework for indigenous language education, planning, and policy McCarty's (2002) framework draws from her work on language revitalization in the United States and compares microlevel and macrolevel possibilities of Indigenous language education in regard to status planning, acquisition/cultivation, and corpus planning (see Table 33.3). McCarty identifies historical and structural constraints (generally originating from institutions of power) that have impacted the micro- and macrolevel possibilities for each area of language planning. McCarty offers several cases including the Diné (Navajo) community of Rough Rock, Arizona, the Hualapai Tribe of Peach Springs, Arizona, and Hawaiian immersion schools. She identifies the constraints throughout instances of status, corpus, and acquisition planning.

- *Status planning.* In the case of Rough Rock, Navajo language was privileged in the "Demonstration School" (the name to which the case school is referred) as opposed to English. This contributed to the microlevel possibility of making Navajo the language of the school despite the assimilative role schools have historically had.
- *Acquisition planning.* In the case of Hawaiian revitalization, although the demand for more Hawaiian schooling programs was met (a microlevel possibility), lack of additional allocation of state funds for these programs made Hawaiian education difficult. This shows that, despite community and parent support of Indigenous language education in schools, these schools still are up against larger obstacles including inadequate school funding.
- *Corpus planning.* Hualapai did not have a "practical writing system" (McCarty 2002: 294) and community members recruited academic linguists to develop one, which contributed to a nationally recognized bilingual/bicultural curriculum (a microlevel possibility). However, non-Indian teachers found the curriculum a threat to students until those objections were eventually worn down.

2.4.3 Ecology of language The ecology of language is a metaphor drawn from environmental work, and stresses that language and the environment interact with each other. Thus the relationship between the two must be considered in research. Carl and Frances Voegelin (1964: 2) explain that with this approach, "one begins not with a particular language but with a particular area, not with selective attention to a few languages but with comprehensive attention to all the languages in the area." Haugen (1972) also offered a definition of language ecology as "the study of interactions between any given language and its environment." He states that "language exists only in the mind of its users" and therefore the ecology can be both psychological and sociological (p. 325). Hornberger (2002: 35) expanded these definitions of ecology of language to offer a three-tiered metaphor in relation to multilingual language policies, where languages "(1) live and evolve in an eco-system along with other languages (*language evolution*); (2) interact with their sociopolitical, economic, and cultural environments (*language environment*); (3) become endangered if there is inadequate environmental support for them *vis-à-vis* other languages in the eco-system (*language endangerment*)." Hornberger contrasts this approach with what has become known as the "one nation-one language" ideology, that is, the idea that a nation-state should be unified by one common language. As this ideology fades under pressure from globalization, space is made for more pluralingual language policies recognizing the feasibility and value of multiple languages in one community or nation-state, such as the Bolivian National Education Reform of 1994.

2.4.4 "The onion" Much LPP research has documented the ways that policies move through language policy actors at different levels. Ricento and Hornberger (1996) offered a metaphor to

Table 33.3 Possibilities and constraints in indigenous language education, planning, and policy development. Reprinted from *Language Policies in Education: Critical Issues* (p. 300), by J. W. Tollefson (Ed.), 2002 Mahwah, NJ: Lawrence Erlbaum. Reprinted with the permission of Taylor and Francis.

Language planning and education processes	Microlevel possibilities	Historical/ structural constraints ⟷	Macrolevel possibilities
Status planning	• Indigenous language as the language of the school, community, and Indigenous nation	• English imperialism • Boarding school legacy • Assimilative role of schools • Political–economic marginalization • Restrictive national policies • Federal paternalism	• Ideological clarification/ commitment • Public valorization for Native languages • Schools as community centers • Development of local leadership • Employment/economical policy development • Local and tribal self-determination
Acquisition/ cultivation planning	• Development of new pedagogies and teaching styles • Heritage language as a first or second language • Enhanced educational achievement and cultural/linguistic pride	• Language attrition • Societal privileging of English • Inadequate school/program funding	• Language revitalization/ maintenance • Identity affirmation • Preparation of indigenous teachers • Development of professional classes/heightening class consciousness • National network of language educators and activists
Corpus planning	• Codification • Elaboration/ modernization • Curriculum planning and development	• Conflicts surrounding authenticity and representation • Pressures for educational "accountability" and standardization	• Creation of new literacies and literatures • Privileging indigenous voices/writers • Strengthened bonds between speakers • Affirmation of sovereignty and local educational control.

explain the various roles different actors have in the language planning process. They consider language planning agents (e.g., policymakers, teachers), levels (e.g., national, institutional, interpersonal), and processes (e.g., legislative) as different "layers" that make up the whole of the LPP "onion." The policy layers – like an onion – permeate each other. Ricento and Hornberger use the onion metaphor to demonstrate the role English language teachers (ELT) have in the language

planning and policy process, powerfully shaping language policies within their own classrooms. For example, despite restrictions placed on native language use in a classroom at a state, district, or federal level, teachers act as policymakers when they allow students to use their native languages to mediate their learning. Hornberger and Ricento suggest that legislation and political processes at the national level are the outer layers of the onion, providing the explicit, overt language policy objectives. These objectives are then interpreted and implemented in institutions (e.g., schools, organized religion, media, libraries, businesses, etc.) as they are often responsible for reproducing the "existing social reality" (Ricento and Hornberger 1996: 416). Classroom practitioners are at the center of the onion. Ricento and Hornberger note that, although practitioners are often thought of as mere implementers of policies designed by others, teachers are as often "primary language policymakers." For instance, teachers might implement an "English-only" policy in an ESL class as a reflection of broader district policy, or might encourage use of minority languages in an English-only school setting.

2.4.5 Funnel Recently Johnson and Johnson (2015) presented a new metaphor, which they describe as a *funnel*. The *funnel* expands on the idea that throughout the LPP process there are implementers, arbitrators, and appropriators. As a policy is funneled through different levels (e.g., from a state to a district to a school), there is an arbitrator (e.g., between the state and the district, and the district and the school) and it is this arbitrator who determines how exactly a policy is implemented. Johnson and Johnson use the example of two school districts in the state of Washington to illustrate the power of the arbitrator. Despite the fact that the districts are under the same state dual language education policy, the structure of the two districts is different; and at each district it is different individuals in different professional roles who are responsible for interpreting the policy to those who then implement the policies differently. These arbitrators are influenced by their own ideologies and beliefs about research, language education, and language learners in addition to pressure from English speaking parents.

The frameworks and metaphors summarized here help articulate how LPP processes interrelate between the so-called "micro" and "macro," and between structure (such as official language planning documents) and agency (such as teachers allowing students to use their native language in a monolingual setting). The frameworks presented by Hornberger and McCarty focus on goals and possibilities of language policy and planning; "the onion" and "the funnel" metaphors, in turn, point to the complexities of language policy across levels of implementation. The ecology metaphor, in turn, firmly locates LPP analysis within the broader social, political, and linguistic environment. These frameworks and metaphors are often used in LPP research, including in the new and developing areas of scholarship and debate highlighted below.

3 Key Areas of Active Scholarship and Debate

The field of language policy has dramatically expanded its reach in recent decades, with respect to both topics of inquiry and methodological approaches. Current scholarship within the field addresses questions such as: How do families determine and enact policies about language in their homes and with what consequences for both child development and language maintenance? What does the linguistic landscape (that is, signage in public spaces) reveal about the value and status of languages in a particular context? How are language policies created, enacted, and contested through social media such as Facebook? How is national immigration policy enacted through language (citizenship) tests and with what impact? As evident here, researchers are engaging with these and many other questions through a wide range of methodological and empirical approaches. Overall, as highlighted above, the field of LPP has moved away from its early emphasis on close examination of official documents outlining top-down national-level policy decisions. Greater attention now paid to how these decisions are understood

by targets (e.g., King and De Fina 2010), reinterpreted by practitioners (e.g., Menken and García 2010), and resisted or transformed by both. This shift in focus has meant greater reliance on close analysis of empirical data, and the development and application of methodologies such as the ethnography of language policy (McCarty 2011), critical discourse analysis, interactional sociolinguistics, and narrative analysis to the study of language policy. Below we highlight several of many active areas of scholarship in the field: family language policy, the language policy of social media, linguistic landscape, Indigenous language revitalization, and ethnography of language policy.

3.1 Family language policy

Family language policy (FLP) bridges studies of child language acquisition and early second language learning and bilingualism on the one hand, with the field of language policy on the other. While the study of bilingual development among children dates back roughly a century (Leopold 1939–49; Ronjat 19131), a defined research focus on family language policy is relatively new. Specifically, FLP examines language policy in relation to language use and language choice within the home among family members (King, Fogle and Logan-Terry 2008). This line of inquiry differs from more psycholinguistically oriented investigations of bilingualism in that, rather than focusing on the child in isolation, the emphasis of FLP is on the balance of languages within the family unit. Thus, FLP addresses child language learning and use as a function of parental language ideologies, decision-making, and strategies concerning languages and literacies, as well as the broader social and cultural context of family life.

The interdisciplinary field of FLP draws from anthropology (and work on language socialization in particular) and sociolinguistics (specifically discourse analysis) while also incorporating traditional psychological approaches to bilingualism. While FLP has coalesced as a (sub)field within the last decade or so (King and Fogle 2013), there have already been several notable shifts. The first of these is increased focus on and intentional inclusion of a broader, more diverse range of family types, languages, and social contexts. While early FLP research tended to document two-parent middle-class homes in which children were acquiring more than one European language (e.g., Métraux 1965; Lanza 1992), current work turns an eye to how these processes play out within minority language and/or nontraditional (e.g., adoptive, single-parent) families in transnational or diasporic contexts (e.g., Canagarajah 2008; Fogle 2012). King (2013), for instance, examines transnational family language policymaking through a longitudinal case study which investigated how linguistic identity was constructed, constrained, and performed by three sisters, aged 1, 12, and 17, within one bilingual transnational Ecuadorian US family. Based on 14 months of weekly home visits (which included participant observation, informal interviews, and audio-recordings of home conversations), King illustrated how each of the three daughters was positioned and positioned herself discursively as a language learner and user, and how ideologies about language and language learning shaped the ways in which identities and family roles were constructed and enacted. These findings sharpen understanding of how widely circulating discourses and ideologies of language, and ideologies of language learning in particular, shape family language practices as well as children's ascribed and prescribed identities within bilingual transnational families.

The second shift in the field of FLP concerns the markedly increased emphasis on the family as a dynamic system, including the importance of child agency and identity choices, both enacted through language (e.g., Gafaranga 2010; Okita 2001). For instance, Fogle's (2012) book examines how Russian-speaking adoptees in three US families actively shape opportunities for language learning and identity construction in everyday interactions. Through close analysis of dinner-table narrative talk, metalinguistic discourse, and code-switching within these families, Fogle demonstrates the varied types of learner agency and shows how language socialization is

collaborative and co-constructed. More specifically, this work documents the ways that children achieve agency through resistance, participation, and negotiation, illustrating the ways in which ways the adoptees transform their homes and communities in transnational contexts.

The third recent shift in the field of FLP is the greater emphasis on trilingualism or multilingualism, with a growing number of studies examining how families manage multiple languages (e.g., Lanza and Svendsen 2007). Curdt-Christiansen (2009), as an example, examined what she termed the "visible" and "invisible" language policies developed among ten immigrant Chinese families in Quebec, Canada. Her analysis focuses on how parents made and implemented decisions with respect to language use and literacy education in three languages: Chinese, English, and French. Her work illustrates the power of "invisible language planning," and in particular parental ideologies related to the sociopolitical and economic markets of each of these languages, in shaping these decisions. Moreover, parental decisions with respect to family language policy were also found to be shaped by Confucian thinking as well as their own immigration experiences and educational backgrounds.

At its core, FLP as an area of research seeks to link parental language policies, strategies, or ideologies with child outcomes in terms of language competencies, identity, or use. This focus work has been framed by extensive and influential writings of Fishman, most notably his book *Reversing Language Shift* (1991), which demonstrated the importance of FLP and intergenerational transmission for language survival. As FLP matures as a field of study, we expect greater focus on the sociocultural factors influencing parents and families as well as sharper attention to the role of children in impacting family language policy processes. With increased attention to the greater number of families participating in migratory and transnational flows, the field of FLP will help us understand the linguistic and social worlds of young children, most of whom – when we look worldwide – are shaped by, and also shape, their multilingual contexts.

3.2 Social media and language policy

A rapidly developing area of research investigates language policy and social media. One line of this work examines the development and negotiation of language policy within social media sites such as Facebook. For instance, Lenihan's (2014) research explores Facebook language policy formation and in particular, the extent to which the site allows for so-called "bottom-up" language policy (Hornberger 2006). Lenihan's analysis demonstrates how the "Translations" app, while appearing to facilitate the development of "language communities" who vote down bad translations as well as report and discuss translations in a specified forum (i.e., bottom-up language policymaking), in fact, operates in a "top-down" manner, with Facebook regularly intervening and adjudicating these discussions, and then authorizing the final translations. Lenihan's work thus shows the ways that LPP is rarely either all "top-down" or completely "bottom-up."

Other work in this vein has examined how language policies are established through social media for the support of Indigenous or minority languages. For instance, Hermes et al. (2016) describe the ways that US Indigenous and endangered language learners and speakers are using social media to promote language learning and provide opportunities for communicative use of the target language.

Another line of this research examines how language policies (of "real world" domains, e.g., of education, nation-states, etc.) are supported, discussed, negotiated, and potentially reframed and resisted in online domains and social media sites. For instance, Rauf-Shier and Farrell (2013) examined the hegemony of English language and its perceived impact on the academic identities of higher education students in Pakistan. To do this, they analyzed the Facebook interactions of higher education students in Pakistan to document their beliefs about how their own academic trajectories have been (re)shaped, and informed by the language education policy in Pakistan.

3.3 Linguistic landscape

A rapidly growing LPP subfield is the study of linguistic landscape. While the first work in this area dates back roughly a decade (e.g., Ben-Rafael et al. 2006), there are currently multiple lines of active research and supporting infrastructure, including regular conferences and a journal (*Linguistic Landscape*, published by John Benjamins, starting in 2015). Studies of linguistic landscape seek to understand the motives, uses, ideologies, varieties, and contestations of "language" (defined broadly) as displayed in public spaces through, for example, signage and advertising as well as public announcements and currency. This research often seeks to uncover inequities and injustices that are largely invisible, routine, accepted, and taken for granted.

Trumper-Hecht (2009), for instance, analyzed the use of Arabic signs in the public spaces of Upper Nazareth in order to shed light on the ways in which national identity is constructed in ethnically and religiously mixed cities in Israel. She found that while the majority of Arab residents of this city (80%) felt that all signs should be in Hebrew *and Arabic*, only 22 percent of Jewish residents shared that opinion. She demonstrates how signage in the city represents the larger political context as well as ideological, emotional, and identity tensions. The language of signs is contested as reflective of the "right to change (in the case of Arabs) and preserve (in the case of Jews) the existing balance of power as well as the very definition of the national identity of the state of Israel" (2009: 250).

Lai (2013), as another example, studied the linguistic landscape of Hong Kong 12 years after the People's Republic of China (PRC) reclaimed sovereignty over the territory. In order to examine the extent to which Hong Kong is a multilingual territory, the influence of the PRC postsovereignty, and the visibility of minority languages in Hong Kong, Lai analyzed 1,160 visual signs displayed in open public spaces in four areas of Hong Kong. She found that Chinese and English far dominated the linguistic landscape of Hong Kong, with Japanese, English, and other languages also existing, but typically for embellishment. Far less visible were other minority languages such as Nepalese and Korean. Lai concludes that

> although English is still highly powerful in the linguistic landscape of Hong Kong as a marker of internationalisation and the local identity, Chinese has become more dominant as it is the first written language of the large majority of the population, and its predominant role being reinforced through the change of sovereignty from Britain to China. (p. 269)

Other work focuses on the impact of globalization and commercialization of spaces. For instance, Lanza and Woldemariam (2014) analyze the use of international brand names and English within the linguistic landscape of Addis Ababa, Ethiopia. While Ethiopia has long been at the margins of the world economy, in the past decade Addis Ababa has experienced a marked uptick in economic growth and corporate investment. As a result, the linguistic landscape of Addis Ababa is increasingly marked by the use of English, not only in general signage but also with respect to international brand names and advertising. Lanza and Woldemariam furthermore document the ways that international brand names and logos are used locally and linked with Ethiopian identity through a process they describe as "cloning." They show how use of both English and international brand names in this particular linguistic landscape is perceived by locals as prestigious, indexing their aspirations toward modernity, as well as geographic and social mobility, in this capital city.

3.4 Indigenous language revitalization

While language shift and language death are not new phenomena, the last three to four decades have been marked by intensified individual, public, and political awareness of the nature and quickening pace of language loss, and of Indigenous language loss in particular (Hermes

et al. 2016; King and Arnal 2016). Hand in hand with this concern has been the development of a vast array of Indigenous language revitalization programs, efforts, and schooling projects. As Grenoble and Whaley (2006: 1) note, "the nature of these initiatives varies as greatly as the languages that are their targets." Some of these initiatives have been national in scope, such as the movement to revitalize Māori in Aotearoa/New Zealand, including the more recent efforts of Te Ipukarea, the National Māori Language Institute, and the International Center for Language Revitalisation at Auckland University of Technology, New Zealand to use technologies as a means for developing high quality materials for language learners (Ka'ai, Moorfield, and O'Laoire 2013). Many other revitalization efforts are smaller in scope, born from the aspirations of a few individuals or communities (e.g., Hermes 2004).

Language revitalization processes and prospects have received extensive attention by language policy researchers, and remain active areas of scholarship and debate. One strand of that work has entailed attempts to describe and define "best practices" in this work: Put simply, "How can language revitalization efforts be successful?" (Grenoble and Whaley 2006: 3), or alternatively, how can less than successful efforts be explained and avoided in the future.

One angle of this work has focused on models of language shift and maintenance. The most well known of these is Fishman's Graded Intergenerational Dislocation Scales (GIDS), outlined in his book, *Reversing language shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages* (1991). Fishman's scale provides a ranking of 1 to 8, estimating the endangerment of the language and vitality of use. For example, a ranking of "1" indicates that the language is widely used in education, media, and at the government level, and a ranking of "8" indicates that the only speakers are elders and it is not being transmitted to younger generations. Levels 1–6 indicate that a language is still being maintained; however, as Dwyer (2011) points out, the scale might not accurately identify languages that are being maintained but are still threatened.

A second area of debate concerns the appropriate roles of academic and scholars within language revitalization. Warner (1999), for instance, discusses ways in which non-Hawaiians appropriate Hawaiian language, culture, identity, and education while silencing Hawaiian voices. He argues that when language work is done by non-natives without collaboration with Hawaiians, Hawaiian identity is further marginalized. Researchers such as Hill and May (2013) assert that non-Indigenous researchers can conduct beneficial research in Indigenous contexts, but only if their work "recognize[s] and addresses the historical imbalances evident in much previous research by non-Indigenous researchers in these contexts" (p. 48). Similarly, McCarty, Skutnabb-Kangas, and Magga (2008) note that language revitalization efforts are often possible due to the work of researchers, but that it is imperative that "the language community itself takes command over all aspects of the use and development of its own language" (p. 306).

An additional line of work examines the broader political context and motivations for language revitalization. For instance, Duchêne and Heller (2007) take a critical or even skeptical approach to language revitalization efforts, seeking to examine not "what makes for language revitalization success," but instead "why and how (and by whom) has language revitalization been framed as a critical goal?" In their words, "rather than simply assuming we must save languages," we should instead be investigating "who benefits and who loses from understanding languages the ways we do, what is at stake for whom, and how and why language serves as a terrain for competition" (p. 11). Within this line of inquiry, Jaffe (2007), for instance, has examined both the causes and consequences of essentializing discourses about language endangerment with respect to the Corsican language. She demonstrates how notions of Corsican language and identity as fixed, ascribed, natural, and unproblematic play out on the island of Corsica. These notions result in, among other things, an emphasis on language-as-code rather than language-as-practice, and views of language communities as homogeneous and static. Other work has taken a more pragmatic stance, examining, for example, the impact of federal education

policy on Indigenous language revitalization efforts in the USA (Cohen and Allen 2013). This work demonstrates the ways that the English-only assessments of federal education policy (and the No Child Left Behind Act in particular) devalue Indigenous social and cultural capital. In particular, in light of the narrowing effect the assessments have had on the curriculum, school–community relations are strained by standards-based reforms.

Most recently, language policy research in this area has put greater attention on the interactional processes of language revitalization, and in particular, what usage patterns promote language learning. For instance, Hermes and King (2013) examined how urban Ojibwe adults utilized computer-based language learning technology with their families at home. Families were observed across two months (e.g., video-taped using the software, interviewed regarding their language use and learning). Findings suggest that technology-based language learning was incorporated into existing family dynamics and was helpful in providing a starting point for learning and language use within established extended networks.

3.5 Ethnography of language policy

As noted above, as the field of language policy has moved away from its early emphasis on analyzing official documents and nation-level decision-making to greater focus on how language polices are implemented, negotiated, and in some cases, resisted, the field has increasingly relied on empirical and interactional data. One current framework for this work which is proving highly productive is the "ethnography of language policy," which can be understood both as a methodological approach grounded in anthropology and also as a perspective for understanding language policy as "processual, dynamic, and in motion" (McCarty 2011: 2). This area of work is driven by questions such as

> What does language education policy "look like" in social practice?; How do policy processes normalize some languages and speakers, and marginalize others?; How are language users and practices "disciplined" or regulated through explicit and implicit policies?; and How do minoritized speech communities exercise agency in the face of oppressive language policies? (p. 4)

A recent example of this scholarship is Nicholas' (2013) ethnographic study of language as cultural practice among contemporary Hopi youth. This study examined how Hopi language shift was experienced across three generations of Hopi: youth, parent, and grandparent. Nicholas finds that while older Hopi perceive a direct link between linguistic competence and cultural identity, younger generations and nonusers of Hopi conceive of cultural identity as part of a process of "practicing culture," hence defining language as a cultural practice. A Hopi way of life, according to Nicholas, is cultivated through participation in the clan-kinship system, the Hopi ritual of planting corn by hand, and song accompanying ritualized performances of social dancing.

This approach has lead to advances in our understanding of the myriad ways in which language policies are variably enacted, understood, negotiated, and transformed across time and space. While macrolevel policy support may be essential, the ethnography of language policy shows "the interpretation and appropriation of top-down language policy is not necessarily predicated on the 'intentions' of the policy" (Hornberger and Johnson 2011: 285). From this vantage point, much of language policy is constructed in interaction rather than written by government ministers or officials. In defining "language policy" as language-regulating modes of human interaction, negotiation, and production (instead of only official acts and documents), the emphasis is taken away from official macrolevel policy texts. Some have questioned this approach. For if we take language policy as not only comprising official acts and documents, but also modes of human interaction and production mediated by relations of power, it becomes reasonable to ask what *isn't* language policy (Hornberger and Johnson 2011: 285)?

4 How LPP Relates to Other Areas/Subdisciplines

Language planning and policy is inherently interdisciplinary. The field is closely linked with, and indeed, methodologically and empirically dependent upon, other disciplines such as history, political science, education, and anthropology. Indeed, many of the scholars cited above would identify themselves first as, for instance, anthropologists or sociologists. LPP routinely utilizes methodology from a number of subdisciplines, including discourse analysis, corpus analysis, and media studies. For example, Fitzsimmons-Doolan (2009) used corpus-based research methods to explore connections between attitudes about immigration and language using texts of Arizona newspapers. Johnson (2010) used critical discourse analysis to analyze bilingual education policy in Philadelphia, Pennsylvania as a way to explore critical connections between the different levels of language policy. And King and De Fina (2010) take up narrative analysis to examine how Latino immigrant women understand US language policy. As the field of LPP has no set, or even preferred, "method," scholars have incorporated and productively combined a wide range of empirical and theoretical approaches to investigate LPP. This has led to the development of a vibrant and expanding area of scholarship.

SUGGESTIONS FOR FURTHER READING

Hult, F. M. and Johnson, D. C. (2015). *Research Methods in Language Policy and Planning: A Practical Guide*. Hoboken, NJ: John Wiley & Sons.

Lo Bianco, J. and Aliani, R. (2013). *Language Planning and Student Experiences: Intention, Rhetoric and Implementation*. Bristol: Multilingual Matters.

Ricento, T. (ed.) (2006). *An Introduction to Language Policy: Theory and Method*. Malden, MA: Blackwell.

Shohamy, E. (2006). *Language Policy: Hidden Agendas and New Approaches*. New York: Routledge.

Vanek, J. (2016). De facto language policy in legislation defining adult basic education in the United States. *Language Policy*, 15, 1, 71–95.

RELEVANT JOURNALS

Current Issues in Language Planning
Language Policy
International Journal of Sociology of Language
Language Problems and Language Planning

EMERGING TRENDS AND RESEARCH QUESTIONS

1. In examining any social context (e.g., a family, a workplace, a classroom), what are the underlying assumptions about which languages should be used for what purposes? What are the working understandings about, for instance, what "language" is, how it is (best) learned, and used? How did these understandings come about and how (and why) are they changing? What language policies are in place, both explicit and implicit?
2. Many researchers are interested in making connections between "big" processes, such as globalization and international immigration, and "small" or "local" decisions, such as deciding what gets translated for whom, or which languages are taught in school. Describing these processes well is challenging, and making claims about causality (what leads to what) is even more difficult. How can connections between the big and small processes be made? What are the perceived ways big and small processes are affected, if not caused, by one another? How have these assumptions been discussed in existing research?
3. Most language policies are not explicit. Choose a particular context (e.g., social media forum)

and look for evidence of implicit policies. How are they determined and enforced? Who has the agency to make them? What happens when an implicit policy is violated? Additionally, what larger power structures (or perhaps explicit policies) exist that might be influencing these practices?

NOTE

1 Ronjat's text is well known for supporting and popularizing Grammont's so-called "one-person-one language" (OPOL) approach as a means to achieve balanced bilingualism, an approach which has been recommended, debated and studied for a hundred years.

REFERENCES

Ben-Rafael, E., Shohamy, E., Hasan Amara, M., and Trumper-Hecht, N. (2006). Linguistic landscape as symbolic construction of the public space: the case of Israel. *International Journal of Multilingualism*, 3, 1, 7–30.

Brown, D. (2013). Legislating language in the name of national unity: an Oklahoma story. *International Journal of Discrimination and the Law*, 13, 1, 4–17.

Canagarajah, A. S. (2008). Language shift and the family: questions from the Sri Lankan Tamil diaspora. *Journal of Sociolinguistics*, 12, 2, 143–76.

Cohen, E. and Allen, A. (2013). Toward an ideal democracy: the impact of standardization policies on the American Indian/Alaska Native community and language revitalization efforts. *Educational Policy*, 27, 5, 743–69.

Cooper, R. L. (1989). *Language Planning and Social Change*. Cambridge: Cambridge University Press.

Curdt-Christiansen, X. L. (2009). Invisible and visible language planning: ideological factors in the family language policy of Chinese immigrant families in Quebec. *Language Policy*, 8, 4, 351–75.

Duchêne, A. and Heller, M. (eds) (2007). *Discourses of Endangerment Ideology and Interest in the Defence of Languages*. London: Continuum.

Dwyer, A. M. (2011). Tools and techniques for endangered-language assessment and revitalization. *Vitality and Viability of Minority Languages*. October 23–24, 2009. New York: Trace Foundation Lecture Series Proceedings. Available online at https://kuscholarworks.ku.edu/dspace/handle/1808/7109

Ferguson, C. A. (1968). Language development. In J. Fishman, C. A. Ferguson, and J. Das Gupta (eds), *Language Problems of Developing Nations* (pp. 27–35). New York: Wiley.

Fettes, M. (1997). Language planning and education. In R. Wodak and D. Corson (eds), *Encyclopedia of Language and Education (Vol. 1): Language Policy and Political Issues in Education* (pp. 13–22). Dordrecht: Kluwer Academic.

Fishman, J. A. (1979). Bilingual education, language planning and English. *English World-Wide*, 1, 1, 11–24.

Fishman, J. A. (1991). *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*. Clevedon: Multilingual Matters.

Fishman, J. A., Ferguson, C. A., and Das Gupta, J. (eds) (1968). *Language Problems of Developing Nations*. New York: Wiley.

Fitzsimmons-Doolan, S. (2009). Is public discourse about language policy really public discourse about immigration?: a corpus-based study. *Language Policy*, 8, 4, 377–402.

Fogle, L. (2012). *Second Language Socialization and Learner Agency: Talk in Three Adoptive Families*. Clevedon: Multilingual Matters.

Gafaranga, J. (2010). Medium request: talking language shift into being. *Language in Society*, 39, 2, 241–70.

Grenoble, L. A. and Whaley, L. J. (2006). *Saving Languages: An Introduction to Language Revitalization*. Cambridge: Cambridge University Press.

Haugen, E. (1959). Planning for a standard language in modern Norway. *Anthropological Linguistics*, 1, 3, 8–21.

Haugen, E. (1972). *The Ecology of Language*. Stanford, CA: Stanford University Press.

Haugen, E. (1983). The implementation of corpus planning: theory and practice. In J. Cobarrubias and J. A. Fishman (eds), *Progress in Language Planning: International Perspectives* (pp. 269–90). Berlin: Mouton de Gruyter.

Hermes, M. (2004). Waadookodaading indigenous language immersion: personal reflections on the gut-wrenching start-up years. In J. F. Ibáñez-Carrasco and E. R. Meiners (eds), *Public Acts: Disruptive Readings on Making Curriculum Public* (pp. 43–54). New York: RoutledgeFalmer.

Hermes, M., Cash, P. C., Donaghy, K., Erb, J., and Penfield, S. (2016). New domains for Indigenous language acquisition and use in the USA and Canada. In S. M. Coronel-Molina and T. McCarty (eds), *Indigenous Language Revitalization in the Americas* (pp. 269–91). New York: Routledge.

Hermes, M. and King, K.A. (2013). Ojibwe language revitalization, multimedia technology, and family language learning. *Language Learning and Technology*, 17, 1, 125–44.

Hill, R. and May, S. (2013). Non-indigenous researchers in indigenous language education: ethical implications. *International Journal of the Sociology of Language*, 219, 47–65.

Hornberger, N. H. (1994). Literacy and language planning. *Language and Education*, 8, 1–2, 75–86.

Hornberger, N. H. (ed.) (1997). *Indigenous Literacies in the Americas: Language Planning from the Bottom up*. Berlin: Mouton de Gruyter.

Hornberger, N. H. (2002). Multilingual language policies and the continua of biliteracy: an ecological approach. *Language Policy*, 1, 1, 27–51.

Hornberger, N. H. (2006). Frameworks and models in language policy and planning. In T. Ricento (ed.), *An Introduction to Language Policy: Theory and Method* (pp. 24–41). Malden, MA: Blackwell.

Hornberger, N. H. and Johnson, D. C. (2011). The ethnography of language policy. In T. McCarty (ed.), *Ethnography and Language Policy* (pp. 273–89). New York: Routledge.

Hornberger, N. H. and King, K. A. (1996). Language revitalisation in the Andes: can the schools reverse language shift? *Journal of Multilingual and Multicultural Development*, 17, 6, 427–41.

Hult, F. M. (2004). Planning for multilingualism and minority language rights in Sweden. *Language Policy*, 3, 2, 181–201.

Jaffe, A. (2007). Discourses of endangerment: contexts and consequences of essentializing discourses. In A. *Duchêne* and M. Heller (eds), *Discourses of Endangerment Ideology and Interest in the Defence of Languages* (pp. 57–75). London: Continuum.

Johnson, D. C. (2010). Implementational and ideological spaces in bilingual education language policy. *International Journal of Bilingual Education and Bilingualism*, 13, 1, 61–79.

Johnson, D. C. (2013). *Language Policy*. New York: Palgrave Macmillan.

Johnson, D. C. and Johnson, E. J. (2015). Power and agency in language policy appropriation. *Language Policy*, 14, 3, 221–43.

Johnson, S. (2005). *Spelling Trouble: Language, Ideology and the Reform of German Orthography*. Clevedon: Multilingual Matters.

Ka'ai, T., Moorfield, J. and O'Laoire, M. (2013). New technologies and pedagogy in revitalization. In M. C. Jones and S. Ogilvie (eds), *Keeping Languages Alive: Documentation, Pedagogy and Revitalization* (pp. 115–27). Cambridge: Cambridge University Press.

Kaplan, R. B. and Baldauf, R. B. (1997). *Language Planning: From Practice to Theory*. Clevedon: Multilingual Matters.

King, K. A. (2001). *Language Revitalization Processes and Prospects: Quichua in the Ecuadorian Andes*. Clevedon: Multilingual Matters.

King, K. A. (2013). A tale of three sisters: language learning and linguistic identity in a transnational family. *International Multilingual Research Journal*, 7, 1, 49–65.

King, K. A. and Arnal, M. (2016). Local and global dimensions of language revitalization in Latin American and the Caribbean. In S. M. Coronel-Molina and T. McCarty (eds), *Indigenous Language Revitalization in the Americas* (pp. 158–78). New York: Routledge.

King, K. A. and De Fina, A. (2010). Language policy and Latina immigrants: an analysis of personal experience and identity in interview talk. *Applied Linguistics*, 31, 5, 651–70.

King, K. A. and Fogle, L. (2013). Family language policy and bilingual parenting. *Language Teaching*, 46, 2, 172–94.

King, K. A., Fogle, L., and Logan-Terry, A. (2008). Family language policy. *Language and Linguistics Compass*, 2, 5, 907–22.

King, K. A. and Haboud, M. (2002). Language planning and policy in Ecuador. *Current Issues in Language Planning*, 3, 4, 359–424.

King, K. A., Schilling-Estes, N., Fogle, L., Lou, J. and Soukup, B. (eds) (2008). *Sustaining Linguistic Diversity: Endangered and Minority Languages and Language Varieties*. Washington, DC: Georgetown University Press.

Kloss, H. (1968). Notes concerning a language–nation typology. In J. Fishman, C. Ferguson, and J. Das Gupta (eds), *Language Problems of Developing Nations* (pp. 69–85). New York: Wiley.

Kloss, H. (1969). *Research Possibilities on Group Bilingualism: A Report*. Quebec: International Center for Research on Bilingualism.

Lai, M. L. (2013). The linguistic landscape of Hong Kong after the change of sovereignty. *International Journal of Multilingualism*, 10, 3, 251–72.

Lanza, E. (1992). Can bilingual two-year-olds code-switch? *Journal of Child Language*, 19, 3, 633–58.

Lanza, E. and Svendsen, B. E. (2007). Tell me who your friends are and I might be able to tell you what language(s) you speak. Social network analysis, multilingualism, and identity. *International Journal of Bilingualism*, 11, 3, 275–300.

Lanza, E. and Woldemariam, H. (2014). Indexing modernity: English and branding in the linguistic landscape of Addis Ababa. *International Journal of Bilingualism*, 18, 5, 491–506.

Lenihan, A. (2014). Investigating language policy in social media: translation practices on Facebook. In P. Seargeant and C. Tagg (eds), *The Language of Social Media: Identity and Community on the Internet* (pp. 208–27). London: Palgrave Macmillan.

Leopold, W. F. (ed.) (1939–49). *Speech Development of a Bilingual Child: A Linguist's Record* (Vols 1–4). Evanston, IL: Northwestern University Press.

Makoni, S. and Pennycook, A. (2005). Disinventing and (re)constituting languages. *Critical Inquiry in Language Studies: An International Journal*, 2, 3, 137–56.

McCarty, T. (2002). Between possibility and constraint: Indigenous language education, planning, and policy in the United States. In J. W. Tollefson (ed.), *Language Policies in Education: Critical Issues* (pp. 285–307). Mahwah, NJ: Erlbaum.

McCarty, T. L. (2011). Introducing ethnography and language policy. In T. L. McCarty (ed.), *Ethnography and Language Policy* (pp. 1–28). New York: Routledge.

McCarty, T., Skutnabb-Kangas, T., and Magga, O. H. (2008). Education for speakers of endangered languages. In B. Spolsky and F. M. Hult (eds), *The Handbook of Educational Linguistics* (pp. 297–312). Malden, MA: Blackwell.

Menken, K. (2008). *English Learners Left Behind: Standardized Testing as Language Policy*. New York: Multilingual Matters.

Menken, K. and García, O. (eds) (2010). *Negotiating Language Policies in Schools: Educators as Policymakers*. Buffalo, NY: Multilingual Matters.

Métraux, R. W. (1965). A study of bilingualism among children of U.S.–French parents. *The French Review*, 38, 5, 650–65.

Nahir, M. (1984). Language planning goals: a classification. *Language Problems and Language Planning*, 8, 3, 294–327.

Neustupny, J. V. (1974). Basic types of treatment of language problems. In J. Fishman (ed.), *Advances in Language Planning* (pp. 37–48). The Hague: Mouton.

Nicholas, S. E. (2013). "Being" Hopi by "living" Hopi. In L. T. Wyman, T. L. McCarty, and S. E. Nicholas (eds), *Indigenous Youth and Multilingualism: Language Identity, Ideology, and Practice in Dynamic Cultural Worlds* (pp. 70–89). New York: Routledge.

Okita, T. (2001). *Invisible Work: Bilingualism, Language Choice and Childrearing in Intermarried Families*. Amsterdam: John Benjamins.

Phillipson, R. (1992). *Linguistic Imperialism*. Oxford: Oxford University Press.

Rabin, C. (1971). A tentative classification of language planning aims. In J. Rubin and B. Jernudd (eds), *Can Language Be Planned? Sociolinguistic Theory and Practice for Developing Nations* (pp. 227–9). Honolulu: East-West Center and University of Hawaii Press.

Rauf-Shier, S. and Farrell, A. (2013, September). *Digital Ethnography of the De-Facto Language Policy of Multilingual Student Groups*. Presentation at Multidisciplinary Approaches in Language Policy and Planning Conference, Calgary, Canada.

Ricento, T. (2000). Historical and theoretical perspectives in language policy and planning. *Journal of Sociolinguistics*, 4, 2, 196–213.

Ricento, T. K. and Hornberger, N. H. (1996). Unpeeling the onion: language planning and policy and the ELT professional. *TESOL Quarterly*, 30, 3, 401–27.

Romaine, S. (2006). Planning for the survival of linguistic diversity. *Language Policy*, 5, 4, 441–73.

Ronjat, J. (1913). *Le développement du langage observé chez un enfant bilingue*. Paris: Champion.

San Miguel, G. (2004). *Contested Policy: The Rise and Fall of Federal Bilingual Education in the United States 1960–2001*. Denton, TX: University of North Texas Press.

Schiffman, H. F. (1996). *Linguistic Culture and Language Policy*. New York: Routledge.

Spolsky, B. (2004). *Language Policy*. Cambridge: Cambridge University Press.

Spolsky, B. (ed.). (2012). *The Cambridge Handbook of Language Policy*. Cambridge: Cambridge University Press.

Stewart, W. (1968) A sociolinguistic typology for describing national multilingualism. In J. Fishman (ed.), *Readings in the Sociology of Language* (pp. 531–45). The Hague: Mouton.

Tollefson, J. W. (1991). *Planning Language, Planning Inequality: Language Policy in the Community*. London: Longman.

Tollefson, J. W. (2000). Policy and ideology in the spread of English. In J. K. Hall and W. G. Eggington (eds), *The Sociopolitics of English Language Teaching* (pp. 7–21). Clevedon: Multilingual Matters.

Tollefson, J. W. (ed.) (2002). *Language Policies in Education: Critical Issues*. Mahwah, NJ: Erlbaum.

Tollefson, J. W. (2006). Critical theory in language policy. In T. Ricento (ed.) *An Introduction to Language Policy: Theory and Method* (pp. 42–59). Malden, MA: Blackwell.

Tollefson, J. W. (ed.) (2013). *Language Policies in Education: Critical Issues*, 2nd edn. Mahwah, NJ: Erlbaum.

Trumper-Hecht, N. (2009) Constructing national identity in mixed cities in Israel: Arabic on signs in the public space of Upper Nazareth. In E. Shohamy and D. Gorter (eds), *Linguistic Landscape: Expanding the Scenery* (pp. 238–52). London: Routledge.

Voegelin, C. F. and Voegelin, F. M. (1964). Languages of the world: Native America, Fascicle One. *Anthropological Linguistics*, 6, 6, 1–149.

Warner, S. L. (1999). Kuleana: the right, responsibility, and authority of indigenous peoples to speak and make decisions for themselves in language and cultural revitalization. *Anthropology and Education Quarterly*, 30, 1, 68–93.

Wolfson, N. and Manes, J. (eds) (1985). *Language of Inequality*, Vol. 36. Berlin: Mouton de Gruyter.

Wong, L. (1999). Authenticity and the revitalization of Hawaiian. *Anthropology and Education Quarterly*, 30, 1, 94–113.

Wright, W. E. (2005). The political spectacle of Arizona's Proposition 203. *Educational Policy*, 19, 5, 662–700.

Zeng, W. (2007) Medium of instruction in secondary education in post-colonial Hong Kong: Why Chinese? Why English? *Working Papers in Educational Linguistics*, 22, 1, 42–56.

Index

AAE, *see* African American English
AAVE, *see* African American Vernacular English
Abailard, P., 99
ablaut, 223
ABSL, *see* Al-Sayyid Bedouin Sign Language
accuracy, 257
Ackerman, F., 231
acoustic level of description, 166
acoustic–phonetic conversion, 336
ACTFL, *see* American Council on the Teaching of Foreign Language
actuation problem, *see* language change
Adamawa-Ubangi languages, 32–33
Adamson, B., 591
Adamson, H. D., 529
adequate perceptual contrast, 164
Adger, D., 529
adverbs, 285–286
  as quantifiers, 293
African American English (AAE), 483, 509, 514, 526, 528–529, 530
African American oral storytelling, 611
African American Vernacular English (AAVE), 520, 528, 532–533, 609, 632
African Eve, 7
  Berber, 31
  Chadic languages, 31
  Cushitic languages, 31
  Egyptian, Ancient, 31
  Semitic languages, 30, 313
agglutination, 222–225
agreement, 217, 218
  subject–object, 380–381
  subject–verb, 129, 379, 380, 386
  verbal, in Garifura, 61–62
Akkadian, 29, 84–85
  see also Sumerian
Albanian, 22–23
Algonquian, 102–103, 110, 313
alignment, 43, 47–48
  ergative, 47, 48, 70, 214, 485
  ergative–absolutive, 51, 485
  nominative–accusative, 48, 51, 70
  object, 47
  range of variation, 51
  subject, 47
allomorphy, 222–223
allophone, 189
Allwright, D., 591
alphabets
  Cyrillic, 85
  roman, 85
  see also writing systems
Al-Sayyid Bedouin Sign Language (ABSL), 391
Altaic languages, 24
Amazonian languages, 70
ambiguity
  elimination, 293
  legal problems, 630
  lexical, 240–241, 293
  structural, 261
  of symbols, 611
  see also language comprehension
ambilingualism, *see* multilingualism
amelioration, 250
American Council on the Teaching of Foreign Language (ACTFL), 593
"Americanist" phonetic symbols, 67
American Sign Language
  acquisition of focus constructions, 385
  chimpanzee, ability of, 10–11
  coordination, 375
  critical period hypothesis, 386
  cross-reference, 374
  effects of delayed acquisition, 386
  iconic vs. symbolic characteristics, 11
  minimal pair, 376, 377
  recursion, 374
  sign structure, 376
  Subject Pronoun Copy, 374, 385
  subordination, 374
  in the courtroom, 639
  word order, 374
  see also sign language
anaphora
  discourse, 450–451
  donkey anaphora, 290
  dynamic, 289–291
  pronoun–indefinite NP, 288
  in structure-building procedures, 434
  and truth-conditional content, 425–426
  as variables, 290
anarthria, 324
Anatolian languages, 23
ancestral language, 7
Andamanese languages, 29
Andersen, H., 307
anthropology, *see* linguistic anthropology
antonomy, 245–247
Anttila, A., 530
aphasia
  Broca's aphasia, 6, 12, 13, 324, 330–331, 337, 387
  effect on semantic processing, 328

The Handbook of Linguistics, Second Edition. Edited by Mark Aronoff and Janie Rees-Miller.

© 2017 John Wiley & Sons, Ltd. Published 2017 by John Wiley & Sons, Ltd.

aphasia (*continued*)
effects of right hemisphere stroke, 334
end-of-sentence comprehension, 329–330
and handedness, 333
semantic roles, problems, 329
sentence comprehension, disorders, 328–330, 336–338
short-term memory impairment, 331
subcortical, 333
syndromes, 324–326, 333, 337
syndromes, correlations with lesions, 334
transcortical motor aphasia, 324
typology, 324–326
Wernicke's aphasia, 324, 325–326, 337, 387
word recognition problems, 335
see also Broca's area; morphology; phrase structure; sign language; word meaning
aphasiology, 323–326
apophony, *see* ablaut
apparent-time hypothesis, 514–515, 522
appropriateness of answers, 148
apraxia, 324
Arabic, 30, 98–99
Cairene, 530
Aramaic, 31, 85
Arawakan languages, 36
Aristoteles, 99
Aronoff, M., 224, 390–391
articulation
aspect, 175–176
clicks, 170
displaced articulations, 173–174
double articulation, 176
fricative, 174, 175
labiodental stricture, 173
linguolabial closure, 173
manner, 169, 186
neutral, 173–174
obstruents, 191
offset, 172
onset, 172, 191, 195
phrasal structure, 172
place of articulation, 168–169, 172–174, 181, 186, 201
primary vs. secondary articulation, 176
resonant, 174–175
stricture, 174–175
topographical aspects, 176
transitional aspects, 177
velar closure, 173
see also sign language; speech
articulatory level of description, 166
ASL, *see* American Sign Language
Asmat, 465
aspect, 215
assimilation, 201
Atlantic-Congo, 32
Atlantic languages, 32
The Atlas of North American English, 531
auditory processing, 398
Austin, J. L., 429
Austin, T. R., 472
Australopithecus afarensis, 5–6
Austro-Asiatic languages, 26
Austronesian languages, 27–28
autism, 401
autonomy of morphology, 219
autonomy of syntax, 153
autosegmental phonology, 202
Axiom of Categoricity, 506

Babel, 100
Babylonian, 97–98
see also Akkadian; Sumerian
Bacon, Roger, 99
Baltic languages, 22
Banfield, A., 473
Bantu languages, 33, 313
Baranowski, M., 532
Basque, 312
Baudouin de Courtenay, Jan, 108
Baumann, R., 475
Bayley, R., 523–524
Beeman, M., 334
behaviorist view of language, 40
Bell, D. M., 593
Bellugi, U., 387
Bender, M. L., 34
Benedict, P. K., 29–30
Benue-Congo languages, 33
Bickerton, D., 13–14
Bilingual Education Act, 655
bilingualism, *see* multilingualism
binding principles, *see* generative grammar
bioprogram, 14
bipedalism, 5–6
Birdsong, D., 565
Bislama, 522
blocking effects, 310
Bloomfield, Leonard, 110, 146
Boas, F., 109–110, 146, 480, 492, 493
Bock, K., 347
Bongaerts, T., 564
Booij, G., 214, 217
Bopp, Franz, 107
borrowing, 106, 310–311, 526–527, 550, 573, 633
see also language change
Bos, H., 385
Bosnian-Croatian-Serbian, 219
Bowker, L., 652
Brahmi, 80

brain
cerebral cortex, 331–332
event-related potential (ERP), 350, 352
imaging technology, 357
inferior parietal lobe, 334
and language, 12–13
lateralization, 333
left inferior frontal gyrus (LIFG), 357
left premotor cortex, 357
location of language processing, 333, 337
location of the lexicon, 337
location of word meaning, 336
N400, 352, 357
organization, 331
P600, 357
pathway, dorsal, 336
pathway, ventral, 336
perisylvian association cortex, 333, 334–335
spatial and linguistic functions in deaf signers, 387
subcortical structures and language processing, 333
supplementary motor area, 332
white-matter tracts, 332
working memory, 466
see also aphasia; Broca's area; language comprehension; Wernicke's area
brain–language coevolution, 6
Brazilian Portuguese, 530
Brazilian Sign Language (LSB), 385
Bresnan, J., 123
Breton, 543
Broca, P., 323–324, 357, 387
Broca's area
location in the brain, 6, 12, 15, 332
location of syntactic processing, 12, 337
in monkeys, 12–13
motor speech planning, 334
see also aphasia; brain
Brugmann, Karl, 105
Buang, 546
Buchstaller, I., 527
bureaucratic language, 638
Burushaski, 29, 312

Calvin, W. H., 13
Cann, R., 7
Caplan, D., 330, 331
Caramazza, A., 329, 337
Carib, 512
Carnap, R., 282, 283, 284
Carstairs-McCarthy, A., 15
case marking, 215
see also alignment
categorial grammar, 284
category–type correspondence, 281
Caucasian languages, 24–25

Cavalli-Sforza, L. L., 7
CBI, *see* communicative language teaching, content-based instruction (CBI)
c-command, 264
CED, *see* Condition On Extraction Domains
Celtic languages, 22
CFG, *see* context-free grammar
Chambers, J. K., 521
Chaucer, G., 299, 300
Cheney, D. L., 8–9
Cherokee, 76
Chibchan languages, 36
Chickasaw, 59
CHILDES corpus, 408
Childs, R., 526
chimpanzees, *see* sign language, American Sign Language
Chomskyan revolution, 111
Chomsky, N.
  on behaviorist psychology, 112
  on caretaker–child interaction, 112
  on competence and performance, 146, 482, 506
  on context-free grammars, 124
  contribution to linguistics, 256–259, 480
  on human language, 141
  on innateness, 112, 486
  on language acquisition, 15, 123, 406, 408, 561
  on language evolution, 4
  on language interpretation, 354–355
  on memory constraints, 353
  on mentalism, 111
  on phrase structure rules, 126–127, 258
  position in the field, 137–138
  on pronominalization, 153
  on the need for language, 345
  on the scope of linguistic theory, 541
  on Universal Grammar, 112
Christianson, K., 347, 356
Chukchi, 214, 215, 217
Chukotko-Kamchatkan languages, 24
Clahsen, H., 360
Clarke, M., 594
cleft sentences, 330
clicks, 488
Clifton, C., 352
CLIL, *see* content and language integrated learning
clitic, 219, 220
CLT, *see* communicative language teaching (CLT)
cochlea, 398
code-switching
  alternative 4M model, 572
  in bilinguals, 572–573
  in Britain, 549, 550
  vs. code-mixing, 574
  equivalence constraint, 572
  free morpheme constraint, 572
  government model, 572
  in Israel, 550
  in Kenya, 549
  metaphorical vs. transactional, 548–549
  motivation, 549
  nonsituational vs. situational, 548–549
  in Norway, 548
  in the US, 551
  see also dialect; multilingualism
Coerts, J., 385
Cognitive Revolution, 345, 346
Cognitive Science, 257
comment, 109
communicative dynamism, 451
communicative function of language, 142–145, 147, 148, 152
communicative language teaching (CLT)
  comprehensible input, 588
  content-based instruction (CBI), 590, 592
  critique, 591–592
  error correction, 589
  Natural Approach, 589
  origins, 587
  task-based instruction (TBI), 590–591, 592
  theory, place of, 594–595
comparative grammar, *see* comparative linguistics
comparative linguistics, 100–106
comparative sociolinguistics, 517
competence, 133, 151
  communicative competence , 482–483, 587
  vs. performance, 111, 121, 146, 208, 506
  transitional competence, 560
Complementizer Constraint, 262, 265, 267
complex predicates, 230
compositionality, 283, 293
  noncompositional interpretation, 293–294
compound(ing), 212–214, 271
  compositionality, 212
  head, 212
  noun incorporation, 213
  see also lexicon; word meaning
Condition On Extraction Domains (CED), 263, 267, 272
conjunction, 281, 373–374
connectionism, 409
constant-rate hypothesis, 528
constraints problem, *see* language change
content and language integrated learning (CLIL), 590
context
  in discourse analysis, 448–449
  sequential context, 452
context-free grammar (CFG), 123–124, 128, 132, 133
  see also generative grammar; semantics
context of situation, 448
conventional implicatures, 426
conversational maxims, 422–426, 431
conversational use of language, 424
Conversation Analysis
  adjacency pair, 452
  and linguistic anthropology, 484
  origins, 454
  repair organization, 452–453
  turn-taking, 438, 452, 454
  see also discourse
converses, 245, 247
cooperative principle, 423
Cooper, R., 292
coordinate conjunction, 132, 137
Coordinate Structure Constraint, 375
Coppieters, R., 565
Coptic, 78–79
Corbett, G., 222
Corsican, 667
Cours de linguistique générale, 108
CPH, *see* critical period hypothesis
cranberry morph, 224
Cree, 83
creoles
  geographic distribution, 37
  vs. pidgins, 13–14
  see also pidgins
critical period hypothesis (CPH)
  in first language acquisition, 386
  in learning to read, 88
  in second language acquisition, 563
  in sign languages, 386
crosslinguistic comparison
  in functional linguistics, 145
  misconceptions, 47
culture, 481
cumulation, *see* fusion
cuneiform, *see* Sumerian
Cutler, A., 359
Czech, 219

Daic languages, 27
D'Arcy, A., 527
data collection
  in anthropology, 481–482
  armchair linguistics, 57, 58
  basic techniques, 58–60
  contribution to theory, 69–70
  data analysis, 67–69
  data publication, 68–69
  difficulties, 63, 65
  in discourse analysis, 446
  displaced speakers, 65
  elicitation, 58–59
  field methods course, 60–61
  finding informants, 62–63
  in generative grammar, 138
  informants, 60, 62–63
  participant observer's paradox, 481–482
  recording devices, 481
  in sociolinguistics, 523–525
  technology, 66–67
  textual analysis, 59, 62
  transcription, 62, 481–482
  see also language samples
Davidson, D., 283–285
Davydova, J., 527
deaf sign language, *see* sign language
DeDe, G., 330, 331
De Guignes, J., 103
de las Brozas, Francisco Sánchez, *see* Sanctius
Dell, G. S., 348
Dene–Caucasian, 30
denotation, 281, 282, 284
  see also sense
derivation, 214, 221, 379
development, language
  and brain size, 6, 14
  population bottleneck, 7
devoicing, 171
  devoiced segments, 178
  initial devoicing, 178
  see also voice
diachronic linguistics
  vs. synchronic linguistics, 302
  see also language change; language variation
dialect
  class dialect, 510
  in code-switching, 548–549
  development from language, 101, 102
  dialect chains, 21
  dialect geography, 525
  dialect maps, 526
  dialect syntax, 528
  and ethnicity, 513
  formation, 305, 308
  vs. language, 21
  multidialectalism, 512, 521
  perception of, 519, 532
  standard dialects, position of, 510
  stigmatization, 532
  variation, 188, 400, 507, 525–526, 529
  see also language variable; language variation
dialectology, 106, 519, 525
Dialect Topography of Canada, 515–516
diglossia, *see* multilingualism
Digueño, 381–382
Dik, S. C., 150
directionality, *see* transformational-generative grammar
discourse
  communicative motivation, 446–453
  context, 448–449
  definition, 445–446
  identity (co-)construction, 456
  ideology construction, 454, 484
  information status, 450
  institutional, 446–447, 454
  institutional knowledge, 456–457, 489
  linguistic identity reconstruction, 454
  macrostructure, 459–450
  participant roles, 455
  plan, 459–450
  reference, 450
  script, 459–450
  see also Conversation Analysis; language ideology; register
discourse analysis
  in forensic linguistics, 636
  place in linguistics, 446, 458
  relation to syntax, 256
  see also discourse; Conversation Analysis
discourse representation theory (DRT), 289, 290
Disjoint Reference Condition (DRC), 264, 267, 273
distinctive features, 109, 167, 199
Distributed Morphology, 225, 226
diversity
  linguistic, 1
Dixon, R. W. M., 29
DNA (deoxyribonucleic acid), 7
Donatus, Aelius, 98, 99
donkey sentence, 292
Drager, K., 525
Dravidian languages, 26
DRC, *see* Disjoint Reference Condition
DRT, *see* discourse representation theory
Dryer, M., 43
duality of patterning, 376
dummy subject, 271
Dutch, 564
Dutch etymologists, 101
Dyirbal, 552
dysarthria, 324

ease of modifiability, 164
ecological robustness, 164
economy of articulatory performance, 164
Edo, 266–268, 274
Educational Testing Services (ETS), 593
Elamite, 29
Elementary and Secondary Education Act, 655
ELF, *see* English as Lingua Franca (ELF)
ellipsis construal, 437–438
Elsewhere Condition, 226
embedding problem, *see* language change
embodied cognition, 357
Emmerik, W., 383
encoded meaning, 419, 422
energeia, 505, 506
English, *passim*
  American English, 177, 187, 596, 243, 304, 305, 520, 530
  Black English, 528
  British English, 171, 174, 177, 180, 188, 243, 596
  Canadian English, 524, 544
  China English, 596
  Early Modern English, 299
  English-based pidgins, 37
  Euro-English, 595, 596
  as a lingua franca, 595–597
  Middle English, 78, 299, 302, 305, 306, 310
  Old English, 78, 299, 302, 304, 305, 306, 468
  Queen's English, 510
  Spanglish, 551
  see also African American English; African American Vernacular English; American Sign Language
entropy, 227
EPP, *see* Extended Projection Principle
ergon, 505, 506
ERP, *see* brain, event-related potential
Eskimo-Aleut languages, 35
Ethnography of Communication, 482

ethnolinguistics, 448, 480
see also anthropology, linguistic anthropology
ETS, *see* Educational Testing Services
etymology, 101, 103–104
E-type pronoun, 290–291
Eurasiatic, 25
evaluation problem, *see* language change
Evans, J., 83
evolutionary approach to grammar, 54
evolutionary biology, *see* origin of language
explicitness, 121, 257
Extended Projection Principle (EPP), 260, 265, 270
extension, 282
see also denotation

Fabb, N., 464, 469, 470–471, 472, 474
Faetar, 523
false friends, *see* translation
Fasold, R., 521–522
feature geometry, 203
feature matrices, 199
feature structure, 126, 134, 135, 136
Felser, C., 360
Ferreira, F., 346, 347, 348
FG, *see* Functional Grammar
fieldwork, *see* data collection
filler-gap dependencies, *see* generative grammar
Final Devoicing, 203–204
Finnish, 181, 225
Finno-Ugric, 23, 313
first-order logic, 280
Firth, J. R., 238, 448, 645
Fischer, J. L., 520
Fishman's Graded Intergenerational Dislocation Scales, 667
Fitch, W. T., 6
Flores, 1
Fludernik, M., 474
fMRI, *see* functional magnetic resonance imaging
focus, 148
Fodor, J. D., 352
Foley, W. A., 142, 146, 275
forensic linguistics
authorship, 631–632
contract disputes, 629–630
copyright infringement, 633
courtroom language, 638–639
defamation, 633–634
eliciting confessions, 637
government transcripts, 635
jury instruction, 637–638
plagiarism, 633
police interrogation, 637
product liability, 628
role of technology, 634–644
trademark infringement, 628–629
voice identification, 630–631
form
vs. function, 46–47, 144
vs. meaning, 223–224, 388, 389
formal deviance, 464
Foucault, M., 445
Foulkes, P., 531
FOXP2 gene, 8
Frazier, L., 350, 355
Frege, G., 280–282, 283, 284
French, 171, 179, 398, 467, 543, 546, 564
Canadian French, 544, 550, 590
early French, 310
French-based creoles, 37
Old French, 650
French Sign Language, 37
Functional Discourse Grammar, 149
Functional Grammar (FG), 39, 149–150, 154
functional linguistics
conservative functionalism, 149
crosslinguistic comparison, 145
explanatory criteria, 152–153
extreme functionalism, 150, 151
vs. formal approaches, 152–154, 275
language function, 142–145
and language typology, 145
pragmatic vs. discourse perspective, 143–144
role semantics and pragmatics, 153
in the United States, 145
functional magnetic resonance imaging (fMRI), 323, 354, 357, 390
functional neuroimaging, 323, 337
functional sentence perspective, 451
function-mapping, 281
function, *see* form
fusion, 223

Garifuna, 61
Gauchat, L., 507, 514
Geeraerts, D., 525
Gelb, I. J., 75, 89
gender, 218
gender roles, 512
generalization
in generative grammar, 111, 121–122
in Structuralism, 110, 111
in the lexicon, 250
generative grammar
antecedent, 127, 128, 129, 137–138
argument structure, 135
binding principles, 127–128, 129, 135
characterization, 259
complement, 127
critique of phrase structure, 133–134
developments, 137–138
explanatory criteria, 152–153
family of theories, 120, 132
feature structures, 126
filler-gap dependencies, 128–130, 135–136
vs. functional approaches, 152–154, 275
island constraints, 131–132, 133, 137, 472
and language function, 603
locality, 127, 129, 135, 431
main principles, 121–123, 151
nonatomic category labels, 125
Principle A, 127, 129, 135
Principle B, 127, 129
Principle C, 356
semantic component, 125
statistical tools, 138
swooping–looping controversy, 130, 137
syntax-orientation, 111
varieties, 132–137
see also context-free grammar; phrase structure rules; transformational grammar
Generative Phonology, 186, 207–208
see also phonology
Genesee, F., 565
genetic affiliation, controversies, 21
genre, 474
German, 185, 405, 409, 528, 551, 552, 560, 647, 650, 651–652
Swiss German, 106, 548
Germanic languages, 22
Geschwind, N., 334
ghoti phenomenon, 77
Gibson, E., 353
Gilliéron, Jules, 105
gloss, 44
glottogonic view, 107
Golston, C., 469
Gothic, 105
government, 217, 218

GQ, *see* quantifiers, generalized quantifier
grammar
  actual vs. possible, 15–16
  definition, 120
  vs. discourse, 604–605
  vs. utterance, 505
grammatical categories, 123, 126
grammaticalization, 113, 214
  see also sign language
grammatical pattern, *see* alignment
grammatical principles
  scope of applicability, 274
grammatical relations, *see* alignment
grammatical structure
  critique, 149–150
grammatical traditions, 97–99
grammatical variable, *see* language variable
grammatical word, 212
grammatology, 90
graphonomy, *see* grammatology
Grassmann's law, 104
Great English Vowel Shift, 516
Great Vowel Shift, 532
Greek, 104
  Ancient Greek, 300
  change in pronoun system, 309
  grammatical tradition, 98
  see also language variation; language change; morphology; writing systems
Greenberg, J. H.
  on an Indo-Pacific grouping, 30
  on competing motivations for implicational universals, 50–51
  founder of modern typology, 113
  on implicational universals, 49
  on Niger-Congo, 31
Grice, H. P., 143, 422–426
Grieve, J., 525
Grimm, Jakob, 104, 313
Grimm's law, 102, 104–105
Grosjean, F., 572
Gumperz, J., 482, 483
Gur languages, 32
Guy, G. R., 530
Gyarmathi, S., 103

Haas, W., 238
Hagoort, P., 359
Hale, K., 69
Halle, M., 306, 469
Halliday, M. A. K., 151, 448
Hamito-Semitic, *see* Afroasiatic
handedness, 6
Hanson, K., 468
The Harvard Dialect Survey, 526
Haugen, E., 656
Hay, J., 525
Hayes, B., 466, 468
Hazen, K., 522
head, *see* phrase structure
Hebrew, 99
Hellenic, 23
Herder, Johann Gottfried von, 107
Herzog, M. I., 301–303, 514, 521
Hickok, G., 337
Hindu, 98
Hittite, 23
Hjelmslev, L., 506
Hmong-Mien, *see* Miao-Yao
Hock, H. H., 307
Hockett, C., 376
Homo erectus, 14
Homo habilis, 14
hominid, 6
homonymy, 212, 241
homophony, 212
Hopi, 668
Hopper, P., 151–152, 473
Hornberger, N. H., 659, 661–663
Hualapai, 661
Humboldt, Wilhelm von, 106, 120, 506
Hungarian, 230–231
Hymes, D., 472–473, 482, 485
hyponomy, 241–242

IA, *see* Item-and-Arrangement theory
iconicity, 53, 54
  see also sign language
Ijo languages, 32
Illič-Svityč, V. M., 25
immersion programs, 553
imperative, 122, 124–125, 128, 133, 134–135
implication, dynamic, 290
implicature, 423, 428, 429
  cancellation, 424
IMT, *see* interlinear morpheme translation
Inclusiveness Condition, 258
incompatibility, 244
Indo-European
  classification, 22–23
  Proto-Germanic, 104
  Proto-Indo-European, 104–105, 306, 312
  recognition as a family, 101–103, 312, 313
  Scythian hypothesis, 101–102, 103
Indo-Iranian languages, 23
Indonesian, 195, 198
inflection
  conjugation, 216
  contextual inflection, 217
  declension, 216
  inherent inflection, 214
  polymorphemic, 225
  principal parts, 227
inner form, 109
innovation, 303, 304–305
  see also language change
intension, 282–283, 291
  see also sense
intensional semantics, 284–287
interlinear morpheme translation (IMT), 44–45
International Association of Forensic Linguistics, 627
International Phonetic Alphabet (IPA), 67, 168, 184, 187
International Sign, 389
interpreting
  consecutive, 646
  décalage, 646
  simultaneous, 646
  syntactic anticipation, 647
intersegmental coordination
  affrication, 180
  anticipatory assimilation, 180
  aspiration, 178–180
  coarticulation and assimilation, 180
  stop release, 180
  voiced aspiration, 179
intertextuality, 494–495
IPA, *see* International Phonetic Alphabet
Irish, 130–131
irony, 473
Iroquoian languages, 35
isiZulu, 488
ISL, *see* Israeli Sign Language
isolated language, *see* language isolate
Israeli Sign Language (ISL), 377, 380, 391
Item-and-Arrangement theory (IA), 221–222
Item-and-Process morphology, *see* Rule Function morphology

Jäger, Andreas, 102
Jakobson, R., 109, 466, 483
James, W., 397, 505–506
Japanese, 24, 194–195, 197–198, 568
jargon, 310
Johnson, D. C., 663
Johnson, E. J., 663
Jones, Sir William, 102–105
Josephus, 101
Junggrammatiker, *see* Neogrammarians

Kanuri, 33
Kanzi (chimp), 11, 404
Karcevskij, Serge, 109
Kātyāyana, 98
Kawahara, S., 470
Khoisan languages, 34, 488
Kilwardby, Robert, 99
Kiparsky, P., 306, 307, 464, 468
Klima, E., 306, 387
Kloss, H., 657–658
knowledge of language and interpretation, 420–421
Kooijman, V., 359
Koran, *see* Qur'ān
Kordofanian, 31
Korean, 24, 83, 190
Kortmann, B., 529
Kraus, Christian Jakob, 103
Kroch, A., 510
Kru languages, 32
Kulick, D., 490
Kuno, S., 149
Kwa languages, 32

Laboratory Phonology, 208
Labov, W.
  on language change, 301–303, 308, 520
  on language variation, 514, 521
  on linguistic stratification, 514
  Martha's Vineyard, 308, 520, 531
  on narrative, 472, 473
  on sociolinguistics, 445, 517, 522
  The Atlas of North American English, 531
Lai, M. L., 666
Lakhota, 143, 146, 275
Langacker, R. 150
Langdon, M., 68
language acquisition (first)
  babbling, 398–399, 400
  behaviorist view, 40
  children vs. adults, 358
  competition between words, 402–403
  confusion about word meaning and its resolution, 402–403
  critical period hypothesis, 386
  early articulation, 398–399
  ergative–absolutive systems, 485
  feedback, processing of, 406
  focus constructions, 385
  guessing word meaning, 401, 403
  infant's preference for mother's speech, 398
  and intelligence, 113
  language awareness in infants, 358–359
  of lexical groups, 407
  mutual exclusivity constraint, 402
  and neural networks, 410
  object concept (Piaget), 399
  overgeneralization, 401, 406–407
  phrasal frequency, 405
  Process Writing, 604
  productivity, role of, 405–406
  psychological views, 409–410
  reading, 621–623
  rule-awareness, 359
  vs. second language acquisition, 562, 604
  semantics, 400–402
  shared reference/eye gaze, 401–402
  social languages, 605
  spelling, 623
  statistical learning, 359
  subcategorization, 359
  syntactic development, 404
  telegraphic stage, 11
  in the womb, 397
  two-word stage, 11, 404, 405
  undergeneralization, 401
  Whole Language, 604
  word-based learning, 404–405
  word-learning, 359, 399–403
  see also American Sign Language; language socialization; sign language; word order
language acquisition (second)
  accent, 564
  access to Universal Grammar, 568–569
  acquisition planning, 658
  approximative systems, 560
  attainable level, 565
  cognitive deficit, 569
  competition model, 567
  contrastive analysis, 559
  critical period hypothesis, 563
  developing links between languages, 567
  error analysis, 559
  vs. first language acquisition, 562, 604
  formal linguistic theories, use of, 558
  fossilization, 560
  generative research approach, 569–570
  importance, 566–567
  influence of age, 563–564
  interference, 559, 573–574
  issues with general linguistic theory, 557–559
  learning sequence, 561–562
  learning stages, 562
  politeness levels, 566–567
  pronunciation, 573
  role in language pedagogy, 588
  spelling, 566
  success levels, 563
  transfer, 559, 566–567
  transitional competence, 560
  vocabulary, 565
  voice onset time, 567
  writing system, 566
  see also code-switching; communicative language teaching; multilingualism
language assessment
  conditioned head turn procedure, 358
  high amplitude sucking paradigm, 358
  intermodal preferential looking paradigm, 358
  in language acquisition, 400, 559
  looking-while-listening procedure, 358
  syntactic comprehension in infants, 404
language change, 42
  actuation problem, 303, 306
  age-related, 514–516
  analogy, 308–310
  borrowing, 310–311
  constraints problem, 301–302, 306
  contact-induced, 491
  contamination, 307
  dating, 516
  embedding problem, 302–303
  evaluation problem, 303, 304–305
  explanations, 308–311
  from innovation, 304–305
  language contact as a source of, 310–311
  lexicon, 304
  Modern American English, 304
  Modern and Ancient Greek, 304
  morphology, 304
  motivation, 307, 308–309
  naturalness, 303, 306
  paths and direction, 302, 303, 308
  in phonological rule systems, 307
  phonology, 304
  predictability, 303
  psychological/cognitive reasons for change, 308–309

language change (*continued*)
recurrent types of change, 305–306
role of speakers, 306
rule-based accounts, problems, 307–308
scope, 303, 307
S-curve, 515–516
social factors, 310
source of sound change, 307
spanning the social and the linguistic, 488
spread of a change, 307
and synchronic variation, 301
transition problem, 302
vowel chain shifts, 532
see also Greek; language variation; Latin; vowels
language comprehension
active filler hypothesis, 352
ambiguity, 349
antilocality effects, 353–354
contribution of the brain hemispheres, 354
distributional probabilities, 352
eye movement, 354, 355, 618
fixation duration, 355, 356, 618
garden-path sentence, 350, 355, 356
grammatical illusion, 356
immediacy principle, 350
interpretation, 356
Late Closure parsing, 350
locality effects, 353
long-distance dependencies, 353–354
memory, 353–354, 466
Minimal Attachment parsing, 350
misinterpretation effects, 356, 360
parsing, vs. comprehension, 354
parsing, one-stage model, 352
parsing, online syntactic, 355
parsing, vs. processing, 354
parsing, two-stage model, 352
predictability, 352–353
processing nonsyntactic information, 351
relative clause, 353
retrieval-activation accounts, 353
role of discourse context, 352
saccade, 355, 618
similarity-based interference accounts, 353
syntactic reanalysis, 350–351, 355
vagueness, 349
visual half-field test, 354
see also reading
language contact, 106, 487, 491–492
language death, 548, 551–552, 553, 666
language diversity, 479, 483
language ecology, 661
language endangerment, 487, 492
see also language shift
Language Flagship Program, 590
language gene, *see* FOXP2 gene
language grouping, 25, 311–313
proposals, 25, 29–30, 34, 36–37
language, human, vs. animal cries, 8–9
language ideology
in education, 489
and (historical) linguistics, 488
language planning, 489
and language socialization, 490
linguistic ideology, 488
linguistic practices operate in social institutions, 489
multilingualism, 490
nationalism and nation-state formation, 489
of monolingualism, 492
role of language in the construction of, 454
see also discourse
language impairment, 410–411
language instinct, 112, 604
language isolate, 312
language learning, *see* language acquisition
language maintenance, 492–493
language pedagogy
audiolingualism, 587, 588
children from lower socioeconomic homes, 608–612
decontextualized audiolingual pattern drill, 589
equity, 607
Exploratory Practice, 592
grammar translation, 587, 588
indigenous language education, 661
K-12 immersion education, 590
medium of instruction, 655–656
methodologies, 585–586
minority children, 608–612
postmethod conditions, 593
PPP procedure, 589
progressive pedagogies, 604
role of grammar, 589
role of teachers, 593–594, 622
situated methodology, 593
skill integration, 588
teacher–pupil communication, 612–613
teachers as mediators of learning, 593
see also communicative language teaching
language planning
acquisition planning, 657, 658, 660, 661
categories, 657–658
corpus planning, 657, 660, 661
definition, 656
ethnography of language policy, 668
family language policy, 664–665
framework of planning goals, 660
historical phases, 658–659
Hong Kong, 590, 655, 666
indigenous language revitalization, 666–668
linguistic landscape, 666
and multilingualism, 665
social media, 665
status planning, 657, 660, 661
top–down vs. bottom–up policies, 656, 665
see also language ideology, language production
language policy, *see* language planning
language preservation, 68–69, 71
language processing, second/foreign language, 360
language production
anticipatory speech errors, 348
in conversation, 348–349
overspecification, 349
planning, 348
radically incremental systems, 347
role of syntactic, semantic, and discourse factors, 347
role of the most conceptually accessible word, 347
syntactic ambiguity, 349
syntactic priming, 348
syntactic structures, 348
underspecification, 349
weakly incremental model, 348
language relatedness, 311–312, 314
language representation, levels, 326
language revitalization, 492–493
see also language planning
language samples
probability samples, 42–43
problem of historical relatedness, 43
randomness, 42–43
required size, 42–44

language shift, 551–552, 666
  Castilianization in Mexico, 492
  Native American languages to English, 492
  Tarascan–Spanish, 491
  as the outcome of language contact, 492
language socialization
  children of migrants, 487
  interactional scaffolding, 486–487
  and language acquisition, 454, 485, 486
  and language ideology, 490
  see also linguistic anthropology
Language Socialization, 454
language standardization, 489
language teaching, *see* language pedagogy
language technology, 138
language universals, 50
  absolute universals, *see* unrestricted universals
  competing motivations model, 50–51
  conceptual space, 51
  constraints on language variation, 50
  economy and iconicity, 53–54
  functional explanations, 52–53
  identification, 50
  implicational universals, 48–49, 109, 113
  innate universals, 40, 359
  semantic maps, 51–52, 53
  and synchronic variation, 302
  and typology, 39, 41
  unrestricted universals, 48
  see also language variation
language variable, 507–509
  age, 514
  classification, 520
  conjugation regularization, 510
  consonant, 531
  ethnicity, 513–514
  final stop deletion, 507
  gender, 511–513
  ING, 520–521, 526
  justification, 521
  morpheme-final consonant cluster simplification, 507
  multiple negative, 510
  nonlinguistic conditions, 508
  phonological, 529, 532
  principle of quantitative modeling, 523
  pronoun plural *youse*, 510
  sex, 511–513
  social class, 509–510
  social network, 511
  voiceless–voiced alternation, 524
  Wolfram's definition, 521
  see also language variation
language variation
  in Ancient Greek, 301
  community support in research, 532–533
  constraints, internal and external, 523
  in contemporary American English, 304
  described in Ancient Rome, 507
  diachronic vs. synchronic, 520
  female-male differences, 512–513, 530–531
  vs. historical variation
  microvariation, 528
  morphological, 526–527
  and parameter theory, 528
  phonetic, 531–532
  phonological, 529–531
  quantitative analysis, 523–525
  rules, internal and external, 523
  sex-exclusive, 512
  in sign languages, 526
  simplification, 508
  in social strata, 508
  source, 523
  standard vs. nonstandard language, 510
  syntactic dialect variation, 528
  syntax, role of, 528
  see also language change; language universals; dialect; language variable; lexical variation; lexicon; Optimality Theory
language vs. linguistics, 119
langue vs. *parole*, 108, 146, 506
  rejection of, 150
Lapp
  relationship with Hungarian and Finnish
laryngealization, 171
larynx
  lowering, 5–6
Latin, 99, 100, 200, 220–221, 224, 306, 307, 309, 467
Lenneberg, E. H., 112, 386, 563
LeSourd, P., 231
Levelt, W., 347
Levin, S. R., 472
lexeme, 211–212, 227, 236
  see also word
lexical access, *see* word meaning
lexical decision task, 328
lexical entry, *see* lexicon
lexical exclusion, 512
lexical integrity, 211
lexical priming, 329–330, 330
  see also aphasia
lexical processing, 336
lexical relatedness, 228–231
lexical semantics
  componential/localistic approach, 239
  conceptual approach, 139
  holistic approach, 238
  structural semantics, 238
lexicon
  in a context-free grammar, 124
  dialect variation, 525
  domain-specific, 253
  expansion, 391
  information structure, 433
  layers, 253
  lexical entry, 212
  lexical fields, 251
  lexical hierarchy, 251
  lexical variation, 208, 525–526
  location in the brain, 335
  mental lexicon, 253–254
  number of lexicons in multilinguals, 573, 574
  place in mental grammar, 523
  role in phonology, 208
  taxonomies, 251
  word families, 252–253
  word fields, 250–252
  see also compounding; language acquisition; language change; language variation; word meaning
lexigram, 11
  see also sign language
Liddell, S. K., 374
Lieberman, P., 5–6
LIFG, *see* brain, left inferior frontal gyrus
like (discourse marker), 300
Lillo-Martin, D., 385
Linebarger, M. C., 329
lingua franca, *see* English
linguistic anthropology
  anthropological view of language, 40
  communicative competence, 482–483
  community, 483
  critique of the Chomskyan paradigm, 486
  vs. ethnolinguistics, 480
  indexicality, 483, 488
  and language origin, 4–7
  see also Conversation Analysis; language planning
linguistic area, 43, 109, 311
Linguistic Atlas of the Middle and South Atlantic States, 525
linguistic relativity, 109, 493
Linguistic Society of Paris, 4

linguistics, science vs. humanities, 106–107
linguistics vs. language, 119
literacy, *see* reading
literary form, 464
literary linguistics, 464, 475
loan translation, 310
Locke, John, 100
logical connectives, 281
logic of conversation, 143
Lotz, J., 468
Love, T., 331
LSB, *see* Brazilian Sign Language (LSB)
Lucy, 5–6
Lwidakho, 549

macaque monkeys, 10
Macaro, E., 594–595
MacEachern, M., 466, 468
MacMahon, B., 471
magnetic resonance imaging (MRI), 12
magnetoencephalography (MEG), 359
Malinowski, B., 448, 481
Mallinson, C., 526
Mande, 32
Mannheim, B., 489
Māori, 181
markedness, 109, 195
Martha's Vineyard, 308, 520
Maslova, E., 43
Mathesius, V. 148
Mayan, 36, 84
McCarthy, R., 327
meaning, 44, 280–294, 309, 607–608
  change, 107
  creation of, 448
  and discourse, 605
  and etymology, 101
  and function, 46
  vs. grammar, 146
  literary texts, 465, 474
  and morphological complexity, 48
  negotiation of meaning, 571, 588, 597
  ordinary vs. definitional meaning, 638
  of a sentence, 279, 329, 419–422
  and sound, 133, 208, 398, 483
  of symbols, 68
  see also form; functional linguistics; lexicon; sign language; word meaning; *signifié*
MEG, *see* magnetoencephalography
Meillet, Antoine, 108, 113
Merge, 258
Meroitic, 34
meronomy, 242
  co-meronomy, 244
Mesrop, 79
metalanguage, 489–490
metaphor, 133, 250, 431–432
metapragmatics, 488
metarepresentation, 473
Methodius, 79
metonomy, 250
Miao-Yao languages, 27
migrants, *see* language socialization
Miller, C., 511–512
Mills, A., 385
Milroy, L., 511
Minimalist Program, 257, 265
mitochondria, 7
Mixe-Zoquean languages, 36
modifiers, 285
Modistae, 99–100
Mohawk, 270–274
molecular genetics, 7–8
Mongolic, 24
Montague, R., 284–285
morph, 221
morpheme, 211
morpheme order, 225
morphological type, 106, 110
  polysynthetic, 271
morphology
  category-shifting morphology, 225
  comprehension problems, 328
  concatenative morphology, 222
  derivational morphology, 227–228
  and language variation, 526–527
  nonconcatenative morphology, 222
  overlap with syntax, 219–221
  processual morphology, 223
  relation to syntax, 256
  subtractive morphology, 223
  in the Ancient Greek tradition, 98
  see also Distributed Morphology; language change; lexicon; Rule Function morphology
morphotactics, 221
Moscow Linguistic Circle, 109
motivated vs. iconic signs, 11
MRI, *see* magnetic resonance imaging
Müller, Friedrich, 107
multilingualism
  ambilingualism, 559
  bilingualism, 559–560
  bilingualism and diglossia, 548
  in Britain, 549
  in Canada, 544
  compound bilingualism, 559
  coordinate bilingualism, 559
  diglossia, 546, 548
  in the European Union, 554
  in Finland, 544
  in France, 543
  incipient bilingualism, 491
  individual and societal, 544–545
  interlingual identifications, 491
  language choice, 545–551
  misconceptions, 542
  in New York City, 545
  organization of linguistic resources, 545–546
  origins, 542–544
  shift toward monolingualism, 551
  in Singapore, 544
  social and situational variables, 546
  in Switzerland, 545
  in Telangana State, India, 545
  see also code-switching; language ideology; language planning
multivariate analysis, 524
Muskogean, 223
mutual intelligibility, 21

Na-Dene languages, 35
Nahali, 29
narrative, 472–473
Nasal Place Assimilation, 203, 205, 206
nationalism, *see* language ideology
national language, 458, 542, 543, 655
  see also language planning
nation-state, 489, 543, 661
natural language interpretation, 425
natural language vs. logical language, 425
natural narratology, 474
nature vs. nurture, 486
Navaho, 403
Neanderthal, 5
Neogrammarians, 105–106, 107, 313
Network Strength Scale, 511
neural network, 407, 409
Newmeyer, F., 153
Nicaraguan Sign Language (NSL), 386–387, 392, 487
Nicholas, S. E., 668
Nichols, J., 149
Nicol, J., 328
Nilo-Saharan languages, 33
Nilotic, 34

Nim Chimsky, 10
N'ko, 83
No Child Left Behind Act, 668
non-narrative evaluative information, 611
nonreflexive pronoun, 127
Northern Cities Shift, 532
Northern Concord rule, 528
Norwegian, 656
Nostratic, 25, 313
NSL, *see* Nicaraguan Sign Language

Occam's Razor, 152
Ochs, E., 485, 486
Odawa (Algonquian), 347–348
official language, 489, 491, 657
  see also language planning
opposites, 247
Optimality Theory, 195–198, 306, 523
  in language variation, 528
  in poetry, 468–469, 470, 471
  variationist models, influence of, 530
origin of language
  archeology, 4–7
  as evolutionary biology, 4
  gestures, 11
  linguistic evidence, 13–16
  nature-versus-convention debate, 98, 100
  philosophical-psychological-typological-evolutionary outlook, 106–107, 109
  single event vs. multiple events, 5
  in the Greek tradition, 98
Otomanguean languages, 35
out-of-Africa scenario, 7, 8

Padden, C. A., 374, 385
Pama-Nyungan languages, 29
Pāṇini, 85, 98
Papiamentu, 527
Papuan languages, 28–29
paradigm, 214, 227
paradigmatic relations, 146
PARLO, *see* Production Affects Reception in Left Only
parole, 146, 506–507
  see also langue
parsing, 436–437
  see also interpretation; language comprehension
participant observation, 481–482
participant observer's paradox, 481
parts of speech, *see* syntactic categories
Patañjali, 98
Pedersen, Holgar, 103
Peirce, C., 483
pejoration, 250
perceptual stability, 164
percolation, 221
performative speech act, 495
periphrasis, 219, 220
PET, *see* positron emission tomography
philology, 90
Phoenician, 79, 86–87
phonation type, 170–173
phone, 189, 377
phoneme, 69, 89, 107, 108, 165, 189, 200, 207, 335, 557, 558, 619, 620, 622, 624
  see also translation
phonetics
  language variation, 531–532
  minimal pair, 188–189, 358
  natural classes, 199
  phonetic equivalence, 165
  phonetic features, 166
  phonetic similarity, 165
  phonetic transcription, 187
  vs. phonology, 201
  sociophonetics, 531
phonetic symbols, *see* International Phonetic Alphabet
phonics, 621, 622
phonology
  adapted phonology, 465
  foot, 467, 468, 469
  laboratory phonology, 208
  mora, 82, 468, 469
  moraic trochee, 465
  vs. phonetics, 201
  role of sign languages, 530
  stress foot, 469
  vs. syntax, 70
  see also language acquisition (second); language change; poetry; sign language
phrase structure (rules)
  cartographic approach, 276
  complement, 127, 131
  constraints, 262–266
  disturbed sentence comprehension, 328–329
  head, 124, 126
  interpretive rule, 261–262
  Japanese, 268–270
  in language production, 348
  macroparametric approach, 275
  mapping problems, 329
  microparametric approach, 275
  recursivity, 120, 124, 256, 372, 373
  relative clause formation, 265
  semantically irreversible sentences, 329, 337
  specifier, 127
  transformational rule, 261–262
  underspecification, 126–127
  vs. word formation, 271
Piaget, J., 399
Pickering, M. J., 351
pidgins
  vs. creoles, 13–14
  geographic distribution, 37
  see also creoles; English; sign language
Pierrehumbert, J., 208
Pilkington, A., 471
Piltdown Man, 6
pitch, 165, 169, 171, 303, 398, 465, 612
Planken, B., 564
Plato, 98
Poeppel, D., 337
poetry
  African American oral storytelling, 611
  alliteration, 470
  cadence, 468–469
  couplet, 465
  difference with other arts, 464
  difrasismo, 471
  enjambment, 471
  epic, 469
  foot, 466, 468
  iambic pentameter, 467, 469
  line, 465
  meter, 466–470, 472, 475
  morae, 463, 466–467
  parallelism, 471
  performance, 494
  rhyme, 463, 470, 475
  rhythm, 466–467
  in sign language, 383–384
  singing, 466
  sound patterning, 470–471
  spoken, 469
  syllable weight, 467
  syntactic freedom, 472
  text-based vs. performance-based, 494
  see also Optimality Theory
Poizner, H., 387
Polish, 203–204
polysemy, 212, 228, 241
Poplack, S., 572
Port Royal, 100
positron emission tomography (PET), 12, 323, 390
possible vs. impossible language, 302
Postal, P. 121–122
poverty of the stimulus argument, 122, 123
pragmatic principles of reasoning, 422–432
Prague School, 108–109, 113, 148, 451
Prakrit, *see* Aramaic

predicates, 404
preposition selection, 129
prescriptive grammar, 119, 121
Preston, D., 521–522
primates, 8–10
priming, *see* lexical priming; semantic priming; language production
principle of multiple causes, 523
Principles-and-Parameters model, 111, 113, 567, 568
Priscian, 98, 99
Production Affects Reception in Left Only (PARLO), 354
productivity, 213
pronominalization, 153
pronoun interpretation, 263
proper names, 286
proposition, 421
  construction, 432–438
  internal structure, 433, 435
  projecting structure, 434
prosody
  alternation, 190
  coda, 192
  compensatory lengthening, 200
  nucleus, 192
  onset vs. rime, 191
  in poetry, 465
  prominence, 190, 468
  Prosodic Word, 190
  sonority, 191, 199
  syllabification, 190, 193
  syllable division, 192–193
  syllable structure, 190–195
  see also articulation; poetry
Proto-Germanic, *see* Indo-European
Proto-Indo-European, *see* Indo-European
proto-language, 13–14, 311
psycholinguistics, 133, 138
  applied vs. developmental, 360–361
  bottom-up vs. top-down processing, 346, 347, 349
  vs. formal linguistics, 346
psychological relevance, 122–123

Quadros, R. M. de, 385
quantifiers
  adverbs, 293
  binding, 288
  dynamic binding, 289
  dynamic existential quantifier (∃{d}), 289, 290
  dynamic universal quantifier (∀{d}), 289
  existential quantifier (∃), 281–282, 289
  generalized quantifier (GQ), 286–287, 291–293
  scope, 288, 293–294
  as second-order functions, 281
  universal quantifier (∀), 281–282, 289
Quechumaran languages, 36
question formation, 269, 272
  see also wh-questions
question movement, *see* question formation
Quichua, 660
Quine, W. V. O., 401
quotatives, 527
Qur'ān, 98–99

Rask, Rasmus, 102, 103–104, 106, 313
Rayner, K., 355
reading
  acquisition, 621–623
  as a psycholinguistic guessing game, 618
  bottom–up processing, 617–618, 621
  cognitive processes involved, 618
  dyslexia, 623–624
  effect on basic cognitive or linguistic abilities, 624
  inadequacy of English spelling, 186–187
  phonemic awareness, 622
  phonics, 621, 622
  and spelling, 623
  spelling-to-sound translation, 620
  top-down processing, 617–618
  whole-language approach, 621
  see also language comprehension; word recognition
reasoning, 424, 426
Received Pronunciation, 171, 510
recursion, 15, 257, 260, 372–373, 473
  see also American Sign Language; phrase structure
reduplication, 223, 464
reference grammars, 44, 45
reference tracking, 144–145
register, *see* social language
Reicher, G. M., 346
relative clause, 46, 130–131, 137, 265, 270, 285, 331, 353–354, 373, 558, 561
Relevance Theory
  cognitive effort and cognitive effect, 427
  cognitive effort and inferential effect, 429
  connectives, 430–431
  context, 427
  explicature, 429, 430
  vs. Gricean pragmatics, 428
  incomplete specific cations, 421
  inferential effects, 427, 428
  interpretation, 430–432
  in literary linguistics, 464–465
  maximizing relevance, 427
  optimal relevance, 428
  recovery, 428–429
  utterance interpretation, 428
reversives, 247
rheme, 148
rhetorical management, 449
rhetorical structures, 449
Riad, T., 469
Ricento, T. K., 661–663
Rickford, A. E., 525, 527
Ritwan-Algonquian, 110
Role and Reference Grammar (RRG), 39, 149
  rules and constraints, 150
  theoretical orientation, 154
Romaine, S., 541
Romance languages, 22, 220
RRG, *see* Role and Reference Grammar
Ruhlen, M., 7, 21
Rule Function morphology, 226–227
Russian, 148–149, 216, 218, 223, 228–229

Saami, 552, 658
Saffran, E. M., 329
Sajnovics, J., 102
Sami, *see* Saami
Samoyedic, 23
samples, *see* language samples
Sanchez, T., 526–527
Sanctius, 100
Sango, 37
Sanskrit, 98, 104–105
Sapir, E.
  biographical sketch, 110
  on language and literature, 464
  on language and thought, 493
  on language variation, 505, 520
  on morphological types, 110
  on the phoneme, 69
  on regularity, 509
  on sociolinguistics, 509
Sapir–Whorf hypothesis, 110
Sarah (chimp), 404
SASS, *see* sign language
satisficing, 356
Saussure, Ferdinand de, 107–108, 146, 506
Savage-Rumbaugh, E. S., 11
Savignon, S. J., 587
Scaliger, Julius Caesar, 100

Schieffelin, B., 485, 486
Schils, E., 564
Schlegel, Friedrich von, 103
Schleicher, August, 105–106, 107
Schuchardt, Hugo, 105–106
Schwartz, M. F., 329
scripts, *see* writing systems
Scythian, 101
Scythian hypothesis, 101–102, 103
Searle, J. R., 143
semantic atoms, 239
semantic memory, 327
semantic priming, 328, 329
semantics, 150
  in context-free grammars, 125
  rise of, 99
  vs. syntax, 132
  of words, 134–135
semiotic modalities of language, 484
sense, 282
  see also denotation
sense relations, 241–249, 251
sentence meaning, 420–421
sentence type, 143, 260
Sequoyah, 76
Serbo-Croat, *see* Bosnian-Croatian-Serbian
Seyfarth, R. M., 8–9
SFG, *see* Systemic Functional Grammar
Shallice, T., 327
Shallow Structure Hypothesis (SSH), 360
signifiant, 108, 146
signifié, 108, 146
sign language
  acquisition, 384–387
  agreement, 380–381, 389
  aphasia, 387–388, 390
  articulation, place, 379
  assimilation, 377
  backwards verbs, 381
  change of sign meaning, 376
  chimpanzees vs. human babies, 11
  classifier predicates, 381–382
  cortex involvement, 332
  cs, left- vs. right-hemisphere damage, 387
  facial expression as intonation, 377
  geographic distribution, 37
  grammaticalization, 391
  iconicity, 376, 380, 382, 390–391
  language birth, 386, 391, 487
  morphological complexity, 382, 390–391
  as a natural language, 371
  phonology, 376–377, 530
  pidginization, 386, 389
  possible signs, 377
  prosody, 377
  sign, semantic, 380
  similarities among sign languages, 388–390
  simultaneity of structure, 378–379
  Size and Shape Classifier (SASS), 382
  typology, 388–390
  variation, 526
  village sign language, 391–392
  see also American Sign Language; poetry
Sign Language of the Netherlands (SLN), 385
sign (semantics), 108, 146, 212, 380
SIL, *see* Summer Institute of Linguistics
Silverstein, M., 483, 488
similitudinal forms, 227
simplification, 486
singing, *see* poetry
Sino-Tibetan, 26–27
Siouan languages, 35
Skinner, B. F., 112
slang, 310
Slavic languages, 22, 148
SLN, *see* Sign Language of the Netherlands
social language, 605
Socrates, 300, 301
Sogdian, 85
Songay, 33
Souleymane Kanté, 83
sound correspondence, as evidence of family relationships, 104
sound inventories, 163, 186–189
sound laws, 105
The Sound Pattern of English (SPE), 202
Southern Vowel Shift, 532
Spanish, 307, 568
SPEAKING, 448
speaking turn, 166
specifier, *see* phrase structure
speculative grammar, 99
speech, 164–165, 181
  affected by sex and gender, 512
  airflow, 160–170
  breathy voice, *see* whispery voice
  creaky voice, 171
  duration, 169, 181
  glottal fry, *see* creaky voice
  murmur, *see* whispery voice
  phonetic quality, 169
  production, 168–181
  segmental classification, 169
  vs. thought, 347
  vocal fry, *see* creaky voice
  whispery voice, 171
  see also articulation; vocal organs
speech act, 142, 143, 147, 429–430, 449, 629
speech activity, *see* speech event
Speech Act Theory, 449
speech community, 483
  see also linguistic anthropology
speech event, 448
speech production, *see* language production
Speelman, D., 525
Sperber, D., 471, 473
SPE, *see The Sound Pattern of English* (SPE), 202
Spolsky, B., 655
Sprachbildung, 107
Sprachbund, *see* linguistic area
Sranan, 37
Sridhar, K. K., 565
Sridhar, S. N., 565
SSH, *see* Shallow Structure Hypothesis
Standard Average European, 40
Staub, A., 352, 354, 355
Steriade, D., 470
Stokoe, W., 376
Stoneking, M., 7
stop, 169, 175
  affricated, 180
  flapped, 177
  glottal, 172
  oral, 179
  stop release, 180
  tapped, 177
structuralism, 107–111
structure dependency, 568
subject, 147
Subject-Auxiliary Inversion, 268
submersion, 552
subordination, 350, 373–374
  see also American Sign Language
Sudanic, 34
Sumerian, 29, 84, 97, 312
Summer Institute of Linguistics (SIL), 70
Sundanese, 202
Supalla, T., 389
suppletion, 214
Survey of California Indian Languages, 68
Swahili, 33, 380
swear words, 512
Swift, K., 511–512
Swinney, D., 328, 330
syllable, 76, 167–168
synchronic linguistics, 302
  vs. diachronic linguistics, 302
  see also language change; language variation

synchronic variation, *see* language variation
syncretism, 216
synonymy, 212, 242–244
syntactic bootstrapping, 359
syntactic categories, 276
syntactic priming, *see* language production
syntagmatic relations, 146
Syriac, 79–80, 85
Systemic Functional Grammar (SFG), 149, 151, 448–449, 451, 603

Tagliamonte, S., 524
tag question, 121–122, 124–125, 133, 134
Taiap, 490
Taiwan, 27
Tarascan, 491
Tarski, A., 283
TBI, *see* communicative language teaching, task-based instruction (TBI)
teaching grammar, 100
Teeter, K., 480
tense, 215, 224, 379
Teodorsson, S.-T., 300
Tewa, 492, 494
text linguistics, *see* discourse
theme, 148, 451–452
Theta Criterion, 260, 265
Thompson, H., 374
Thompson, S. A., 473
Tibetan, 85–86
Timbuktu, 33
Tocharian, 23
Tok Pisin, 490, 546
topic, 109, 147, 148, 152, 451–452
topicalization, 375, 434–435, 437, 439
touristese, 14
transformational-generative grammar, 111, 124–125
  directionality, 133
  incorporation of semantic component, 132–133
  language learnability, 123, 133
  see also context-free grammar; generative grammar; phrase structure
transition problem, *see* language change
translation
  computer-assisted translation, 652
  correctness, 648
  definition, 645–646
  false friends, 650
  handling untranslatable concepts, 652
  interlingual false friends, 649–650
  machine translation, 652
  phonemic, 648–649
  semantic factors, 651–652
  synchronic intralingual false friends, 650–651
transposition, 225, 229, 230
tree diagram, 123, 124, 133, 136
Tristram, H. L. C., 470
Trousdale, G., 529
Trubetzkoy, N. S., 109
Trumper-Hecht, N., 666
truth-conditions, 283, 426
  context-dependency, 426, 431
  truth-conditional semantics, 425
truth-functional connectives, 283
truth value, 281
Tucanoan languages, 36
Tungusic, 24
Turkic, 24
Turkish, 203
typology
  categories, 106
  empiricist approach, 40
  functional classification, 46–47
  vs. generative and functionalist approaches, 39, 113–114
  questionnaires, 45
  rationalist approach, 40
  resources, 44
  of sign languages, 388–390
  statistical techniques, 51
  typological approach to language, 46–47
  typological diversity, 42
  see also crosslinguistic comparison; functional linguistics; language universals; Universal Grammar

uniformitarian principle, 301
Universal Grammar, 49–50, 112, 122, 375, 408
  counter-evidence, 359
  headedness parameter, 528
  rise of, 99–100, 133
  and second language acquisition, 567, 568–570
  see also typology
universals, *see* language universals
usage-based grammar, 150
Uto-Aztecan languages, 35–36, 110

Vai, 76–77
Van Valin, R., 142, 146, 275
VARBRUL, 525, 529
Varro, 507, 517
Vedas, 98
verbal art, 493, 494
  see also poetry
vernacular studies, 517
Verner, Karl, 105
Verner's law, 105
verse, *see* poetry
vervet monkeys, 8
vocabulary, *see* lexicon
vocal call systems, 8–9
vocal organs, 5–6, 10, 170, 168, 170
Voegelin, C. F., 480, 661
voice (anthropological), 494–495, 605
voice (phonological), 168, 170
  assimilation of, 201
  as a distinctive feature, 199
  phonetic feature vs. distinctive feature, 167
  voiced aspiration, 179
  voicing alternation, 203
  see also devoicing; speech
voice (syntactic), 147
Voiced Stop Deletion, 205–206
vowels, 186
  acoustic characterization, 165
  loss of unaccented vowels, 309
  vowel harmony, 203
  vowel length, 181
Vygotsky, L. S., 608

Wackernagel position, 219
Waletzky, J., 472, 473
Walmatjari, 523
Warren, R., 347
Warren, T., 353
Warrington, E. K., 327
Washoe (chimp), 10, 404
Wasow, T., 527
Waters, A., 589–590
Waters, G. S., 330, 331
Webb, R., 389
Weinreich, U., 301–303, 514, 521, 559, 567
Welsh, 552
Wernicke's area, 13, 332, 334
Wette, R., 593
wh-questions, 128–129, 132, 261, 267, 330
White, L., 565
Whitney, W. D., 520
Whorf, B., 110, 403
Wilson, A. L., 7
Wilson, D., 471, 473
Wolfram, W., 521, 529, 530
word
  definition, 235–236
  demarcation, 236
  effect of context, 249
  grammatical properties, 236
  literal vs. nonliteral readings, 249–250

meaning extension, 249
nonce word, 249
as a partial structure projection, 433
possible word, 185
semantic properties, 237–238
signed vs. spoken, 375–376, 380
see also ambiguity; language acquisition; lexeme; lexicon; sense relations; sign language
word deafness, 324
word meaning
contextual variability, 239–241
disturbance, 327–328
and the inferior parietal lobe, 334
influence of lesions, 336
and lexical access, 335–336, 337
meaning extension, 249–250
metaphorical extension, 212, 250
nature of, 336
predicting, 346
retroactivation, 336
sense relations, 241–249
vs. sentence meaning, 419
speaker's intentions, 419
see also brain; compounding; language acquisition; lexicon
word order
acquisition, 404
Al-Sayyid Bedouin Sign Language, 391
explaining word-order differences, 275
Mohawk, 270–271
object-first languages, 70
in second language acquisition, 560–561
Sign Language of the Netherlands (SLN), 385
in Slavic languages, 451
see also alignment; American Sign Language; phrase structure; question formation; universals
word recognition (in aphasics), 335
connectionist models, 620
lexical and nonlexical routes, 620
and writing system, 619–620
see also reading
word-superiority effect, 346
Wray, A., 14
writing, 4, 631
writing systems and scripts
abjad, 79–80
abugida, 80–81
alphabetic, 77–79, 619
alphabet vs. other systems, 89
diffusion, 85–87
Ethiopic, 81
grapheme, 88
hieroglyphs, 31, 86
hiragana, 82
hPags-pa, 86
Indic, 80–81
inventions, 76, 83
kanji, 82
katakana, 82
vs. language, 88–89
logographic, 76, 619
logosyllabary, 76
morphosyllabary, 81–82
origins, 84–85
pictograms, 82
syllabary, 76–77
see also language acquisition (second); reading; word recognition
Wulfila, 79
Wundt, Wilhelm, 113
Wurm, S., 28

X-bar theory, 127

Yabem, 546
yes–no question, 429
Yip, M., 464

Zaenen, A., 130
Zapotec, 492
Zeng, W., 656
zero copulas, 41
ZISA, 560
Zulu, *see* isiZulu
Zunshine, L., 473
Zurif, E. B., 328, 329, 337
Zwicky, A., 527