

**The Use of Articles in Inner and Outer Circle Varieties of
English: A Comparative Corpus-based Study**

Ridwan Wahid

A thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

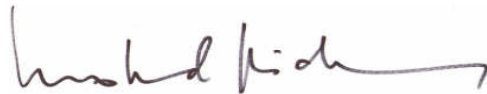
School of Languages and Linguistics
University of New South Wales

2009

Originality Statement

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Signed

A handwritten signature in black ink, appearing to read 'Mushad Kichu', written in a cursive style.

Date:

31 August 2009

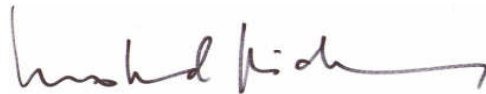
Copyright Statement

'I hereby grant the UNSW or its agents the right to achieve and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the three hundred and fifty word abstract of my thesis in Dissertation Abstract International.

I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material.'

Signed

A handwritten signature in dark ink, appearing to read 'Lushad Kich', written over a light grey rectangular background.

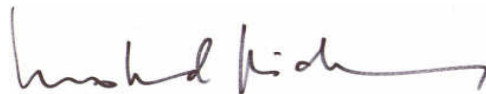
Date

31 August 2009

Authenticity Statement

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed

A handwritten signature in dark ink, appearing to read 'Lushad Kich', written over a light grey rectangular background.

Date

31 August 2009

Acknowledgements

I owe a number of people my gratitude for their contribution to the completion of this thesis. First of all I thank Peter Collins, my supervisor, for his guidance, support and encouragement. I am grateful for the interest that he showed in the study and for the openness with which he approached the study of Outer Circle varieties with me.

I also thank Mengistu Amberber for his encouragement and interest in the study and Debra Aarons for the support that she gave, especially in my progress reviews. Both of them also generously allowed me to sit in their syntax classes. I am grateful to Kayoko Evon for the excellent administrative support that she provided and to Margarita Grebennikov for allowing me to use the writing laboratory as I needed. My thanks also go to Vincent Ooi with whom I had many hours of insightful discussion on corpus linguistics and varieties of English when, once again, our paths crossed at UNSW.

For enabling access to its database, I thank Macquarie University's corpus linguistics research group. I also extend my appreciation to Gerald Nelson for giving me permission to use several of the ICE components.

I must thank my wife, Suhaila, for her unending support and for the faith that she placed in me. I owe the completion of this thesis to her. I am also grateful to our daughter, Reihana, who continually, yet cheerfully, lifted the weight of academia off my shoulders. To my family and Suhaila's family in Malaysia, I extend my heartfelt

thanks for their continued support in spite of the distance. Finally I thank our Malaysian friends in Sydney, Khairil and Hidayah, for their assistance in double-checking the cluster frequencies.

This thesis is dedicated to the loving memory of my mother, Noriah Ali. *Biarpun kecil, ini masih untuk Mak. Terima kasih yang tidak terhingga – al-Fatihah.*

Abstract

Articles continue to be one of the most puzzling aspects of English grammar. This thesis investigates article use variation in two groups of English varieties: Inner Circle and Outer Circle. The first group, comprising British English, American English, Australian English and New Zealand English, represents English varieties in their traditional contexts. The second group, consisting of Singapore English, Indian English, Philippine English and Kenyan English, represents varieties which are deeply entrenched in their communities due to colonial roots. The study is premised on the naturalness of corpus data. Three corpus-based analyses were conducted to investigate articles and their frequency, collocation and usage types.

The frequency analysis revealed that articles are idiosyncratically used by the varieties with no clear divide between the two groups. Text types, on the other hand, emerged as a more powerful determinant of the observed frequencies. It was found that fewer articles were used in spoken language. In the collocational analysis, the varieties were found to favour different complementation patterns for two-word and multiple-word clusters. The Inner Circle varieties, however, shared more complementation patterns among them, demonstrating greater stability in terms of variation. Finally the analysis of usage types discovered ‘marked’ or untypical patterns of use in the Outer Circle group’s structural and situational definites. A number of untypical usages of first-mention/quantitative *a/an* were also found in this group. In the same analysis, a case study of ‘house’ uncovered a few marked uses of \emptyset , which signalled the null form but not the zero article.

Although the collective findings in this study indicate that article use in both groups of varieties has more similarities than differences, the latter leave more impression on observers. These differences can be attributed to such diverse factors as text types, socio-cultural content of collocates, specificity, recategorisation of countness, indiscriminate rule application and ungrammaticality. Theoretically, however, the variation can be captured by the application of Chesterman's three composite meanings of definiteness – locatability, inclusiveness and extensivity – and their permutations. As an instance of nativisation, article use in the Outer Circle is limited in its diffusion. The observed innovations rarely creep into more formal registers.

Table of contents

Originality Statement	i
Copyright Statement	ii
Authenticity Statement	ii
Acknowledgements	iii
Abstract	v
List of Tables	xiv
List of Figures	xiv
List of Concordances	xviii
Chapter 1 Introduction	1
1.1 Article use: second language acquisition vs. nativisation	2
1.2 What constitutes the article system?	6
1.3 Research questions	21
1.4 Organisation of the chapters	22
Chapter 2 Articles and Definiteness	23
2.1 Definite and indefinite articles	24
2.1.1 The definite article <i>the</i>	24
2.1.2 The indefinite article <i>a/an</i>	28
2.1.3 The zero article	30
2.1.4 The generic article	32
2.2 What is the meaning of definiteness?	37

2.2.1	Uniqueness	38
2.2.2	Familiarity	43
2.2.3	Inclusiveness	47
2.2.4	Identifiability	48
2.3	Reference and specificity	51
2.4	Genericity	54
2.5	Unifying accounts of definiteness	56
2.5.1	Abbott (2004)	56
2.5.2	Huebner (1983)	59
2.5.3	Lyons (1999)	63
2.5.4	Chesterman (1991/2005)	65
2.5.5	Preliminary conclusion	67
2.6	Some previous SLA studies on articles	68
2.7	Summary	70
Chapter 3	Nativisation	72
3.1	Language change and variation	72
3.1.1	Innovation	76
3.1.2	Diffusion/spread	80
3.1.3	Nativisation as a process of language change and variation	85
3.2	World Englishes	91
3.3	Describing varieties	93
3.3.1	British English	94

3.3.2	American English	95
3.3.3	Australian English	97
3.3.4	New Zealand English	98
3.3.5	Indian English	99
3.3.6	Singaporean English	100
3.3.7	Kenyan English	102
3.3.8	Philippine English	104
3.4	Concluding remarks	105
Chapter 4	Research Design	106
4.1	Why a corpus-based approach?	106
4.2	The International Corpus of English (ICE)	107
4.2.1	Santa Barbara Corpus of Spoken American English	109
4.2.2	Freiburg-Brown Corpus of Written American English	110
4.2.3	Corpus of Contemporary American English and Brigham Young University – British National Corpus	110
4.3	Levels of analysis	111
Chapter 5	Frequency and Variation in Article Use	113
5.1	Article deletion and insertion	113
5.2	The study	118
5.3	Procedures	119
5.4	Results: <i>the</i>	121

5.5	Discussion: <i>the</i>	128
5.6	Results: <i>a/an</i>	149
5.7	Discussion: <i>a/an</i>	152
5.8	Conclusions	153
Chapter 6	The Collocations of Articles	155
6.1	A study of the collocational patterns of articles	155
6.1.1	Collocates of <i>the</i> : some general observations	156
6.2	Right and left collocates of <i>the</i>	187
6.2.1	R-1 collocates: nouns	187
6.2.2	R-1 collocates: adjectives	208
6.2.3	L-1 collocates of <i>the</i>	210
6.2.4	Longer-span collocates of <i>the</i>	220
	6.2.4.1 Clusters of <i>the</i> in the Inner Circle varieties	229
	6.2.4.2 Clusters of <i>the</i> in the antipodean varieties	234
	6.2.4.3 Clusters of <i>the</i> in the individual varieties	239
6.3	The collocates of <i>a/an</i>	240
6.3.1	R-1 collocates of <i>a/an</i>	241
	6.3.1.1 R-1 collocates <i>a/an</i> : general nouns	247
6.3.2	L-1 collocates of <i>a/an</i>	253
6.3.3	Longer-span collocates of <i>a/an</i>	254

	6.3.3.1 Clusters of <i>a/an</i> in the Inner Circle varieties	257
	6.3.3.2 Clusters of <i>a/an</i> in the antipodean varieties	268
	6.3.3.3 Clusters of <i>a/an</i> in the individual varieties	274
6.4	Conclusions	283
Chapter 7	The Usage Types of Articles	287
7.1	Revisiting the usage types of the definite article	287
	7.1.1 Hawkins' usage types of the definite article	288
	7.1.2 Prince's taxonomy of givenness	292
7.2	Annotation schemes	294
	7.2.1 Developing an annotation scheme for <i>the</i>	294
	7.2.2 Developing an annotation scheme for <i>a/an</i>	303
	7.2.2 Analysing \emptyset	304
	7.2.4 Data and procedures	305
	7.2.5 Sample cases and problematic cases for the definite article	307
7.3	Results: <i>the</i>	316
	7.3.1 Textual and associative anaphoric <i>the</i>	324
	7.3.2 Structural <i>the</i>	324
	7.3.3 Situational <i>the</i>	330
	7.3.3.1 The situational definites and the cline of acceptability	333
	7.3.4 Factive, generic and idiomatic <i>the</i>	338

7.4	Results: <i>a/an</i>	338
	7.4.1 Idiomatic <i>a/an</i>	345
	7.4.2 First-mention/quantitative <i>a/an</i>	346
	7.4.3 Factive and generic <i>a/an</i>	350
7.4.4	Other untypical characteristics of <i>a/an</i> in Outer Circle varieties	350
7.5	Results: \emptyset	351
7.6	Conclusions	362
Chapter 8	Conclusion	367
8.1	A synthesis	367
	8.1.1 <i>The</i>	368
	8.1.2 <i>A/an</i>	371
	8.1.3 \emptyset	372
	8.1.4 All articles	372
8.2	Research questions	374
	8.2.1 Article use: similarities and differences between Inner and Outer Circle varieties	374
	8.2.1.1 Frequency differences	375
	8.2.1.2 Collocational differences	375
	8.2.1.3 Usage type differences	376
	8.2.2 Explaining the differences	376
	8.2.3 Nativisation vs. SLA	380
	8.2.3.1 L1 influence	380
	8.2.3.2 Proficiency	381

	8.2.3.3 Specificity	382
	8.2.3.4 Nativisation	382
8.3	Limitations of the study and recommendations for future research	384
	Bibliography	385
	Appendix	403

List of Tables

Table 1.1	Use of articles with count and noncount nouns	18
Table 2.1	The uses of <i>the</i> , <i>a/an</i> , zero and generic articles	35
Table 2.2	Huebner's environments for the appearance of <i>the</i> , <i>a/an</i> and zero (adapted from Thomas, 1989)	62
Table 2.3	Distribution of articles across Chesterman's oppositions constituting definiteness (1991/2005: 68)	67
Table 3.1	Criteria catalogue for ESL varieties (Mollin, 2007: 173)	90
Table 4.1	ICE text categories	108
Table 5.1	Profiles of L1s found with IndE, SgE, KenE and PhilE	117
Table 5.2	Frequencies of <i>the</i> across varieties and text types	122
Table 5.3	Residuals in a chi-square relationship between a variety and its text types for <i>the</i>	125
Table 5.4	Residuals in a chi-square relationship between a text type and varieties for <i>the</i>	127
Table 5.5	Frequencies of <i>the</i> in Inner vs. Outer Circle varieties	128
Table 5.6	Frequencies of <i>a/an</i> across varieties and text type	151
Table 6.1	General and specific nouns among the top 1000 R-1 collocates of <i>the</i> (spoken)	191
Table 6.2	General and specific nouns among the top 1000 R-1 collocates of <i>the</i> (written)	192
Table 6.3	General and specific nouns in 100-intervals (spoken)	196
Table 6.4	General and specific nouns in 100-intervals (written)	200
Table 6.5	Adjectives with high socio-cultural content among the top 1000 most frequent R-1 collocates of <i>the</i>	208
Table 6.6	Shared clusters according to Inner and Outer Circle varieties	225

Table 6.7	General nouns as R-1 collocates of <i>a/an</i> among top 500 words in all varieties	249
Table 6.8	Clusters of <i>a/an</i> in Inner and Outer Circle varieties	256
Table 6.9	Clusters of <i>a/an</i> in each individual variety vs. others	276
Table 7.1	Correspondence between Hawkins', Prince's and the present study's categories	302
Table 7.2	Tokens of <i>the</i> according to varieties and categories	317
Table 7.3	Percentages of each category of <i>the</i> according to varieties	319
Table 7.4	Tokens of <i>a/an</i> according to varieties and categories	340
Table 7.5	Percentages of each category of <i>a/an</i> according to varieties	342
Table 7.6	<i>House</i> according to determination types and varieties	351

List of Figures

Figure 1.1	Phrase syntax	10
Figure 1.2	The definite article in the DP structure	11
Figure 1.3	Distribution of forms in nominal syntax (from Klinge, 2006: 59)	13
Figure 3.1	Sources and processes leading to postcolonial varieties (from Schneider, 2007: 100)	78
Figure 5.1	Percentages of underuse of <i>the</i> in the varieties	131
Figure 5.2	Percentages of overuse of <i>the</i> in the varieties	133
Figure 5.3	Bar chart showing percentages of variation and proportions of overuse and underuse of the definite article in each variety	139
Figure 5.4	A comparison of selected collocates of <i>the</i> in NZE and and BrE	144
Figure 5.5	A comparison of the most frequent R-1 collocate of <i>the</i> in W2A in AmE and BrE	146
Figure 6.1	Most frequent noun collocates of <i>the</i> at R-1 position in all varieties (except US where noted)	156
Figure 6.2	Most frequent adjectival collocates of <i>the</i> at R-1 position in all varieties	161
Figure 6.3	The gradient between proper names and definite NPs (adapted from Quirk et al., 1985: 294)	180
Figure 6.4	Ooi's Concentric Circles Model (1997; 2001)	206
Figure 6.5	Most frequent prepositional collocates of <i>the</i> at L-1 position in all varieties	210
Figure 6.6	Most frequent non-prepositional collocates of <i>the</i> at L-1 position in all varieties	214
Figure 6.7	Clusters found in top 200 in all varieties	221
Figure 6.8	Most frequent L-1 collocates of <i>a/an</i> in all varieties	241

Figure 6.9	Top 12 L-1 collocates of <i>a/an</i>	253
Figure 6.10	Clusters shared by all varieties among their top 200	255
Figure 7.1	Correspondence between Liu and Gleason's and Hawkins' categories	295
Figure 7.2	Categories for <i>a/an</i>	304
Figure 7.3	A series of bar charts showing the percentages of each category of <i>the</i> according to varieties	320
Figure 7.4	The cline of acceptability of the (situational) definite article	337
Figure 7.5	A series of bar charts showing the percentages of each category of <i>a/an</i> according to varieties	343
Figure 8.1	A diagrammatic representation of findings and their interpretation	373

List of Concordances

Concordance 6.1	50 lines of <i>round the</i> from Inner and Outer Circle varieties	211
Concordance 6.2	10 lines of <i>also the</i> from Inner and Outer Circle varieties	216
Concordance 6.3	10 lines of <i>this the</i> from Inner and Outer Circle varieties	217
Concordance 6.4	10 selected lines of <i>put the</i> from Inner and outer circle varieties	219
Concordance 6.5	5 selected lines of ‘preposition + <i>behind the</i> ’ in Inner and Outer Circle varieties	220
Concordance 6.6	10 lines of <i>a while</i> from Inner and Outer Circle varieties	244
Concordance 6.7	11 lines of <i>a nice</i> in ICE-Ken and 11 selected lines of <i>a nice</i> from all other varieties	245
Concordance 6.8	7 lines of <i>a moment</i> in ICE-Ken and 7 selected lines of <i>a moment</i> from all other varieties	246
Concordance 6.9	17 lines of <i>I’ve got a</i> in Outer Circle varieties and 20 selected lines of <i>I’ve got a</i> from Inner Circle varieties	258
Concordance 6.10	17 lines of <i>we’ve got a</i> and <i>he’s got a</i> in inner circle varieties and 20 selected lines of <i>we’ve got a</i> and <i>he’s got a</i> in Outer Circle varieties	260
Concordance 6.11	24 lines of <i>a range of</i> in Outer Circle varieties and 25 selected lines of <i>a range of</i> from Inner Circle varieties	261
Concordance 6.12	14 lines of <i>and it’s a</i> in Outer Circle varieties and 14 selected lines of <i>and it’s a</i> from Inner Circle varieties	263
Concordance 6.13	47 lines of <i>be/was a bit</i> in Outer Circle varieties 50 selected lines of <i>be/was a bit</i> in Inner Circle varieties	264

Concordance 6.14	22 lines of <i>a bit more</i> in Outer Circle varieties and 25 lines of <i>a bit more</i> in Inner Circle varieties	267
Concordance 6.15	37 lines of <i>have a look at</i> in the Outer Circle varieties and ICE-GB and 40 selected lines of <i>have a look at</i> from ICE-AUS and ICE-NZ	269
Concordance 6.16	14 lines of <i>for a couple of</i> in the Outer Circle varieties, 9 lines of <i>for a couple of</i> in ICE-GB and 39 lines of <i>for a couple of</i> in ICE-AUS and ICE-NZ	271
Concordance 6.17	7 lines of <i>hell of a</i> in the Outer Circle varieties, 7 lines of <i>hell of a</i> in ICE-GB and 24 lines of <i>hell of a</i> in ICE-AUS and ICE-NZ	273

Chapter 1 Introduction

This study has an applied outlook but is firmly descriptive in its approach. Its main objective is to investigate an aspect of the English language that has long puzzled linguists and lay people alike. This is the use of articles. The issue that originally motivated the study is also a professional one for some – learners (and teachers) of English across time and place are beleaguered with the question of how to use articles properly. Indeed it has been the subject of numerous studies and many a proposal has been offered as *the* solution. But, as we shall see, this is far from being definitive.

In the present study, I re-examine this issue by paying attention to two increasingly important areas in contemporary linguistics: world Englishes and corpus linguistics. The first one is a theoretical platform that provides a more objective way of thinking about the English used by speakers outside its traditional contexts. The second one is more methodological in nature and complements the first by promoting the use of naturally occurring data that have often been de-emphasised in many research contexts.

This study investigates the use of articles in eight varieties of English. These include British English, American English, Australian English, New Zealand English, or following Kachru's term (1985), varieties from *Inner Circle* countries. Also included are Singaporean English, Indian English, Philippine English and Kenyan English, or *Outer Circle* varieties.

The next two sections will outline other main issues in the research of articles that are relevant to this study. These are (1) article use in relation to second language acquisition versus nativisation in 1.1 and (2) the article system in English in 1.2. In 1.3 the research questions are presented. The final section, 1.4, will describe the organisation of the chapters that constitute this thesis.

1.1 Article use: second language acquisition vs. nativisation

It is well-known that the use of articles by non-native speakers of English shows a large amount of variability. But where is the variability found? It is not difficult to think of a context in which this is especially true. The language classroom readily comes to mind. As pointed out above, learners of English are known to struggle with article use. It is regarded as one of the most difficult grammatical aspects to master by non-native speakers (Berry, 1991; Master, 1990; Robertson, 2000). Consequently there exists a large body of research on articles from the second language acquisition (SLA) perspective. Its main agenda is to ascertain the causes of the difficulties that are faced by the learners. Some SLA researchers explore the syntactic causes of this phenomenon (e.g. Hawkins, 2001; Trenkic, 2007 *inter alia*). Others focus on its semantic aspect (e.g. Huebner, 1983; Ionin and Wexler, 2003). Still other researchers explore the issue by integrating the two perspectives exclusively (e.g. Ionin et al., 2004) or with others such as discourse (e.g. Jarvis, 2002), sociolinguistics (e.g. Sharma, 2005) and language pedagogy (e.g. Goto-Butler, 2002; Master, 1997). The focus on learners has naturally led to a scrutiny of the learners' interlanguage. The notions of nativeness and error are widely used to contextualise the learners' usages that depart from

accepted norms. Unfortunately this is where the problem lies. While these constructs have certainly facilitated the effort to understand the learners' problems, their usefulness comes into doubt when dealing with the use of articles by speakers in non-pedagogical contexts. If our contention is that non-native speakers *in general* use articles variably, we need to re-consider the relevance of the findings that are specific to second language (L2) learners. Many of the speakers in the non-learning contexts are in fact those of the Outer Circle varieties mentioned above. We will return to these speakers and their varieties below.

Other than its heavy reliance on the constructs of nativeness and error, SLA research has also prominently featured the use of elicited data. Although elicitation is certainly a good method to gather data, it is not without shortcomings. Primary among these is the lack of naturalness in the production. Such output could compromise the validity of the findings (see chapter four for a fuller description of corpus data). Again, when considering speakers of Outer Circle varieties, such a concern must be properly addressed. One simple reason is these speakers do not always use English with native speakers (in fact, many of them rarely do) and therefore, naturalness becomes an important factor to consider. This brings us to the relevance of the native speaker to the enquiry being made here.

There is an increasingly accepted view that the nativeness construct should be re-examined given the spread of English around the world. Support for this view comes most visibly from scholars working in a largely sociolinguistic area that has come to be

known as *world Englishes*. Many historical events and the dynamics of present-day social, political and economic affairs around the world have turned English into a uniquely ‘world’ language (Schneider, 2007). Native English speakers come into contact not only with people from different cultures but also languages that are very distantly or not at all related to it. It is no wonder then that the resulting non-native varieties of English are found to exhibit linguistic variations that range from being minimal to entirely novel. Acceptance of the variation, too, varies widely. Thus although the global and widespread use of English is accepted, it is not uncommon that these non-native varieties are regarded with disfavour (Bhatt, 2001). Understandably at the heart of the resistance is the desire to preserve the perceived purity of certain native varieties. However due to the common rejection of the newer varieties’ status – and the indisputable fact that English is, and will continue to be, used outside its native contexts for a host of different reasons – linguists have questioned the notion of nativeness that is frequently associated with arguments seeking to relegate the non-native varieties to a lesser level. Kachru (1982) proposed the term ‘world Englishes’ to refer to such varieties so as to better accommodate the diverse forms that English has now evolved into. Put another way, English used in Jamaica is as valid as that used in Australia although there obviously exist differences in the linguistic features between the two.

The spread of English is explained by Kachru through the now-classic concentric circle model. The Inner Circle represents the traditional contexts of English (i.e. the United States, Britain, Australia, Canada and New Zealand). The Outer Circle refers to non-native contexts of English in which it has reached the status of an official, second or

additional language. These include countries like India, Singapore, Malaysia, Nigeria, etc. The outermost circle is the expanding circle, representing contexts where English is used primarily as a foreign language. These include most European countries other than the UK and Ireland, countries like Japan, Saudi Arabia, South American countries, China, Korea, etc. Although this model is sometimes criticised for not taking into account the rapid globalisation of English (e.g. Bruthiaux, 2003), the essential division regarding the uses to which English is put in the contexts identified remains relevant (see 3.1 and 3.2).

While it has become clear now that one has to look beyond the native speaker for the provision of norms, this cannot be equated with a call to disregard the findings of SLA studies. It has to be emphasised that the pedagogical context in which many of these studies were embedded requires that a certain linguistic standard be upheld or attained and this, until recently, has often been that of a native variety. On a more critical level, however, what the findings of such studies help point out is that acceptance of a native variety as a standard is always a controversial issue. Once a native variety's standard is accepted as one's own to achieve, usages deviating from the norms of that variety are necessarily seen as unwanted and treated as errors. However if we factor out variables such as standards, prestige, acceptability etc. from the issue, we will find that the variation patterns are various degrees of nativisation in various forms.

It is the Outer Circle varieties of English that are found to demonstrate the highest degree of nativisation (Lowenberg, 1986). Out of the need to suit local needs and

because of sustained use, innovations become necessary and are often codified (Bangbose, 1998). Studies that exemplify such innovations in such varieties include Bangbose (1995) on Nigerian English, Kachru (1983) on Indian English and Platt and Weber (1980) on Malaysian and Singaporean English.

So the question now is: is the use of articles, including those usage patterns that deviate from normal and/or accepted norms, a phenomenon of SLA or nativisation? Because the focus of the study is on the varieties in Inner and Outer Circle countries, the answer must be the latter. However accepting that article use is subjected to nativisation entails the re-examination of the findings in SLA research from the perspective of world Englishes. It is a reasonable assumption that the findings will differ either quantitatively or qualitatively (or both) from those of SLA. This is what this thesis sets out to do.

1.2 What constitutes the article system?

In this section I intend to clarify the term ‘article’ as used throughout this thesis. The morphemes that I refer to as articles are the definite article *the*, the indefinite article *a/an* and the phonologically null article, zero or \emptyset . Following Chesterman (1991) it will be shown in chapter 2 that the phonologically null article can be further divided into two – zero and null. Some linguists accept unstressed *some* as in (1) as an article:

1. I'd like some water please.

For example, Chesterman argues that “[t]he obvious reason for including *some* is its perfectly complementary distribution *vis-à-vis* (typical) *a*, occurring before plural and

mass nouns, but not singular count nouns” (p. 45). However, as Huddleston and Pullum et al. (2002: 358-359) show, *some* is – first and foremost – quantitative:

- 2a. *Some men* are still in that house (stressed, proportional, i.e. “not all” implicature for count nouns);
- 2b. *Some water* is left in that glass (stressed, proportional, i.e. “not all” implicature for noncount nouns);
- 3a. There were *some men* in that house (unstressed, non-proportional for count nouns);
- 3b. There was *some water* in that glass (unstressed, non-proportional for non count nouns).

Some in (2) and (3) can also express *multal* quantification in the sense that both sets of men or water in the four sentences are fewer/less in quantity than those indicated by certain other quantifiers such as *many*, *a lot*, *much*, *a great deal*, etc. (Huddleston and Pullum, 2002: 365). This strengthens its quantitative interpretation. On *some* vs. zero, Quirk et al. note that “[t]he distinction we have to draw, then, is between CATEGORIAL meaning of zero, and the QUANTITATIVE meaning of *some*” (1985: 275). It will be shown below that the main contrast between definiteness and indefiniteness as indicated by *the* vs. *a* (originally a quantifier) is ‘which’ and not ‘how much/many’. Correspondingly, I will exclude *some* from the study and focus on zero only for indefinite plural and non-count noun phrases NPs.

In Modern English, articles precede the head of an NP and are part of a class of such pre-head words called determiners. Because the ‘determining’ function can also be

carried out by words not belonging to this group (e.g. a genitive NP in *Ray's glasses*), the specific name for such words, including articles, is *determinative* (Huddleston and Pullum, 2002: 365). There are three classes of determiners: predeterminers, determiners and peripheral modifiers. For our purposes, suffice it to say that articles are determiners.¹ This position is illustrated in:

4. Not *even* *half* *the* *land* will be yours.
 peripheral modifier predeterminer determiner head

A question that some linguists ask is why articles are classed as articles at all. It has been pointed out that *the* and *a* are not syntactically equal as determiners (Klinge, 2006), necessarily leading to the different meanings between them (Trenkic, 2006). Their inequality can complicate some analyses of their patterns. Here are some cases to illustrate:

(i) *A/an* can sometimes alternate as a numeral (as in *for a second*) in which case it can be replaced by *one*. However when expressing genericity, the replacement cannot take place – to express an equivalent meaning of ‘*The red panda is a beautiful creature*’, *a* not *one* must be used, as in ‘*A red panda is a beautiful creature*’;

(ii) *The*, on the other hand, is neutral as to number marking and can be used in singular, plural and noncount NPs;

(iii) *A/an* occurs with singular count nouns due to its origin as the numeral one. The plural counterpart to *a/an* is either *some* or zero. But it can occur with mass nouns in

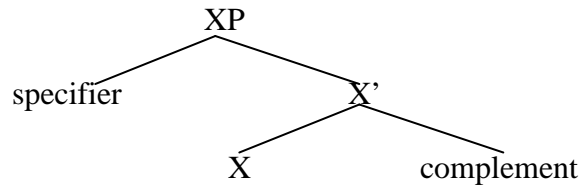
¹ Quirk et al (1985: 261) classify articles as ‘central determiners’ in relation to predeterminers and postdeterminers. The latter “follow predeterminers or central determiners (if such determiners are present)...but they precede any adjectives and other premodifying items”. In Huddleston and Pullum et al. (2002), the central determiners appear to be the determinatives themselves.

certain contexts, as in ‘I have *a good knowledge* of computer programming’. However, no plural equivalent is possible i.e. * ‘I have *some/∅ good knowledges* of computer programming’.

It appears that any coherent description of articles must not only acknowledge that they are derived from very different sources in the grammar of the NP, but also that they are grouped together as members of a single class because they serve a common purpose. I will discuss at some length here how articles can be viewed as grammatical morphemes which are exponents of the semantic field of (non)identifiability. In other words they fulfill different syntactic roles although they serve the same semantic function in the NP. This explains both their common and incompatible properties and ultimately, their complementary distribution. This discussion will set the stage for the analysis of *the*, *a/an* and zero in the chapters ahead.

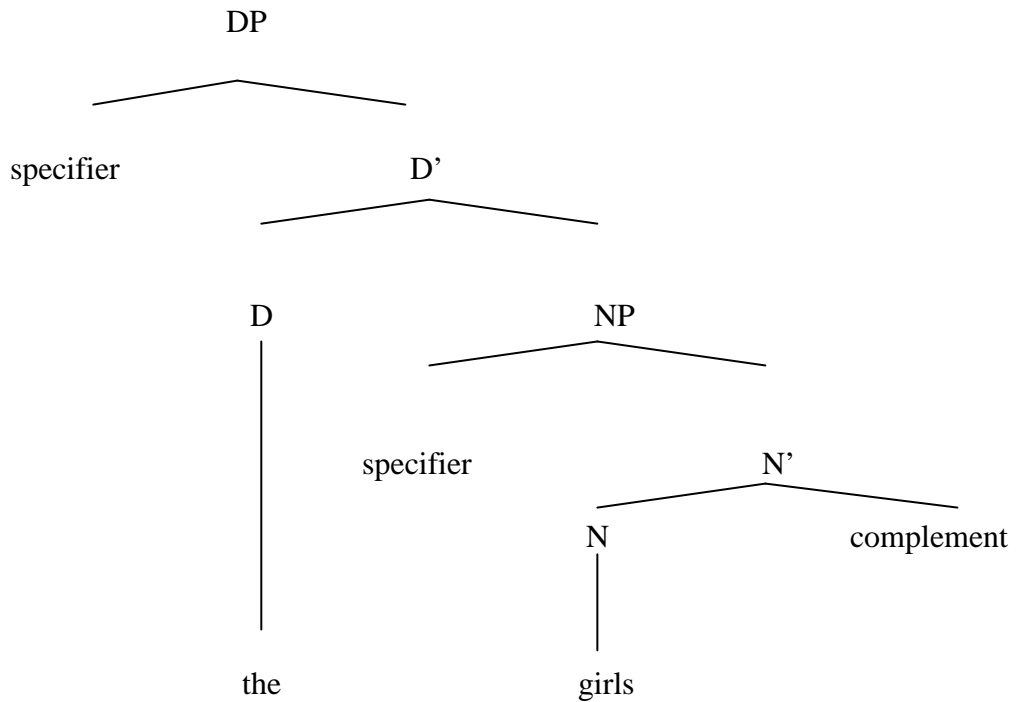
In generative syntax the phrase is seen as a projection of a head. In Figure 1.1, X is the head of the projection, culminating in XP, its maximal projection. This schema is known as X-bar theory. A generally accepted view until recently is that the noun phrase (NP) is the maximal projection of the head N. On this view, the article occupies the specifier position and is therefore on the same level with (or ‘sister’ to) N’.

Figure 1.1: Phrase syntax



In line with the increased attention given to functional categories such as tenses, modality, agreement, etc. in clause structures, there has been a shift from the above view in regard to the NP. In particular it has been argued that despite the fact that articles, demonstrative pronouns, possessive pronouns and genitives are in complementary distribution (thus occupying a single ‘specifier’ slot), it is misleading to conceive of them as constituting a single category. Haegeman and Guéron (1999: 409-411), for example, point out that articles form a closed class and contribute only marginally to the descriptive content of the NP. In addition, in some languages (e.g. Italian) the definite article can co-occur with a possessive pronoun. Observations such as these led to the conception of the DP, a ‘functional’ (as opposed to ‘lexical’) projection with the determiner (such as articles) as head. Thus, NP, although remains projected from N, is now within the higher DP and a complement of the head D. This can be seen in figure 1.2:

Figure 1.2: The definite article in the DP structure



The structure above will more adequately explain certain movements in structures such as ‘How severe a situation is this?’ (cf. Huddleston and Pullum, 2002: 328-332 on the structures of external dependents, predeterminers and peripheral modifiers without determiner as head).

Syntactically a few constructions are shown to demonstrate sensitivity to definiteness. This is known as the definiteness effect. I will describe here only the most common construction that shows this effect. This is the existential sentence or the *there*-sentence. To illustrate, consider the following:

- 5a. There is a rabbit in the fridge.
- 5b. ? There is the rabbit in the fridge.
- 6a. There were some maids in the kitchen.

6b. ? There were the maids in the kitchen.

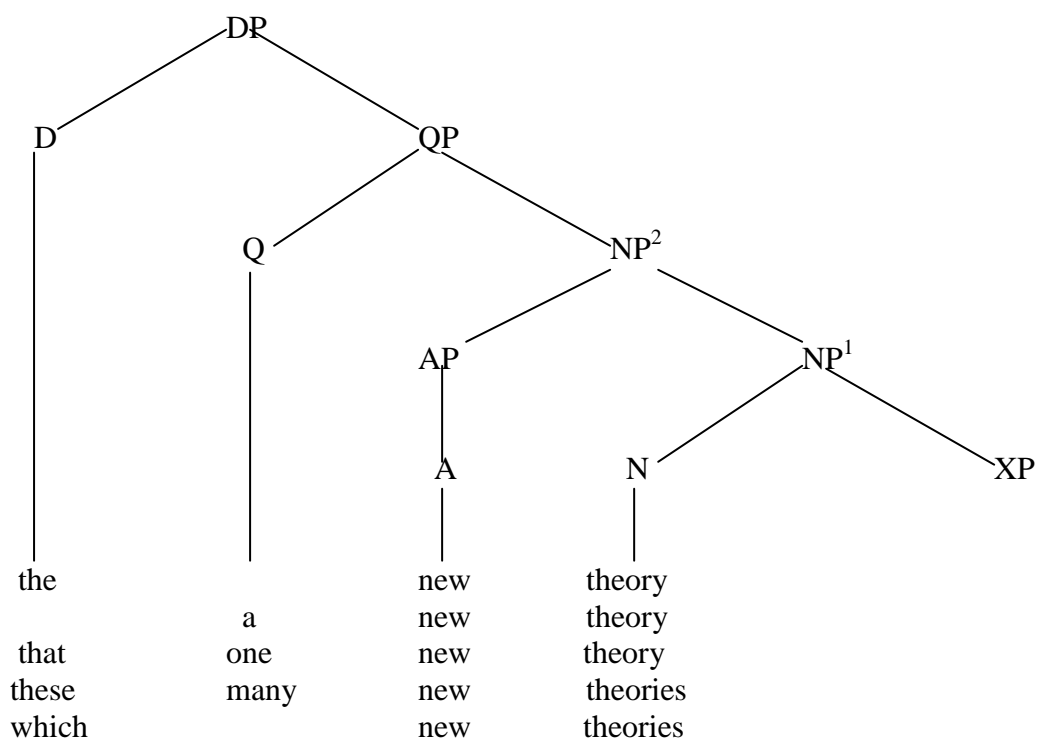
The *there*-construction is usually used as a diagnostic to determine whether an NP² is definite or not. However, in uses where the NP serves to “fulfil some role or purpose” (Abbott 2006: 5), (5b) and (6b) become felicitous. For instance, (5b) could be used as a response to the question ‘What can I add to the stew to make it more delicious than this?’ and (6b) for ‘Weren’t there people in the house when the roof came down?’ In each of these uses, the prepositional phrase modifies the definite NP and does not serve to give a locative meaning to the whole clause.

The above DP structure, however, does not make a distinction between definite and indefinite articles. Dixon points out that “[a]most every grammarian of Modern English groups *the* and *a/an* together, either as part of a major word class (generally adjective, but pronoun and preposition have also been suggested....) or as a separate class” (2006:31). On the contrary, as implied by Dixon’s statement, this is not a position adopted by all linguists. Arguing that *the*, *this* and *that* should be unitarily treated on the one hand, and *a* with other cardinal quantifiers on the other, Klinge claims that “[t]he facts of complementary distribution with other quantifiers and the fact that it has retained its singular quantification strongly suggests that [*a/an*] distributes in paradigmatic contrast with lower quantifiers” (2006:57).

² I follow the practice of many other writers of using the term ‘NP’ to refer to *noun phrase*, while acknowledging that both the determiner and nominal form a DP.

As for *the*, it is known that it does not enter into paradigmatic contrast with *one*, *two*, *three*, *many*, etc. (e.g. ‘the one book’). It is also commonly assumed that *the* is derived from the singular demonstrative *þē* in Old English (Klinge, 2006:58). On the basis of these observations Klinge postulates that *the*, *this* and *that* are ultimately derived from the Old English non-proximal demonstrative paradigm. This, Klinge argues, attests to a level of shared semantics between these three determiners, and not with the so-called indefinite article (ibid.). In DP terms this can be illustrated as:

Figure 1.3: Distribution of forms in nominal syntax (from Klinge 2006:59)



Indeed as pointed out by Trenkic (2004:1422), several researchers have noted the different qualities of the definite and indefinite articles beyond their obvious

oppositeness. Chesterman, for example, states that both articles represent “qualitatively different concepts” (1991/2005:1).

If this structural account is taken to be true for the articles, what justification can be provided for a study that attempts to analyse them as equals? Or what is the basis for collectively treating them as markers of (in)definiteness? A straightforward answer would be like one provided by Klinge where he states, “*the* has definiteness as part of its semantic consequences so it cannot co-occur with [*a/an*] for the simple reason that [*a/an*] has indefiniteness as part of its semantic consequences” (2006:60).

While this answer is adequate, it does not explain how *both* articles can be used to express genericity (see 2.4), or the fact that in languages such as French the two articles can co-occur as in:

de	la	viande
PART/INDEF	DEF	meat

To regard the indefinite article as being on a par with the definite article, we need to examine its evolution from the numeral one.

The evolutionary model described here is Heine’s. Heine claims that the stages contained therein can be interpreted both synchronically and diachronically. Synchronically it is an “implicational scale”, i.e. an indefinite article at a particular stage may subsume all the properties of earlier stages but not vice-versa. Diachronically

the model represents the evolution on a time scale (1997:71-76). The model is summarised below:

1. Stage I: The numeral

An item for 'one' is available as a numeral

e.g. ni-na gari moja

(Swahili)

I- have car one

'I have one car.'

2. Stage II: The presentative marker

The article is used to introduce "a new participant presumed to be unknown to the hearer and this participant is taken up as definite in subsequent discourse" (p.72)

e.g. Zhyl da byl odin starik ...

(Russian)

lived PARTICLE was one old. Man

'Once upon time there was an old man ...'

3. Stage III: The specific marker

The article is now used to mark a discourse referent known to speaker but not hearer.

The referent "may or may not be taken up in subsequent discourse" (p.73)

e.g. ba hena ish- xad etmol ve- hitxil le- daber ve- hu

(Street Hebrew)

came here man-one yesterday and started to- talk and- he

‘A man came in yesterday and started talking and he ...’

4. Stage IV: The non-specific marker

The article is able to introduce a discourse referent known neither to speaker nor hearer

e.g. a. Buy me a newspaper, please!

b. Draw a dog!

5. Stage V: The generalised marker

The article is able to introduce all types of nouns – with few exceptions – including singular, plural and mass

e.g. Un día ven- ían un- o- s hombres

(Spanish)

one day come- 3PL.PRET.IMPERF one- M- PL men

‘One day there came some men ...’

Notice that in English the indefinite article is able to carry out the functions associated with all stages except the last one. This fact requires us to reconsider the designation of ‘one’ as an equal of *a/an* at all times. Consider:

7a. To awaken her, the princess must be kissed by one prince.

7b. To awaken her, the princess must be kissed by a prince.

The typical reading for (7a) is one and specific prince (stage III). On the other hand, what the indefinite article does to the nominal ‘prince’ in (7b) is to mark it as any

prince (stage IV).³ Clearly *one* is not capable of expressing this meaning and this must be taken to mean that *a/an* has grammaticalised into a marker that is beyond a pure numeral (see Heine 1997:80). It can be said then that the cardinal meaning of *a/an* is superseded by the speech acts that make use of the indefinite article. Sometimes this means one, and perhaps specificity too (e.g. He has *a wand*). But in other contexts it can signal non-specificity (e.g. He wants *a wife*). A full syntactic treatment of this interpretation is, however, outside the scope of this study (see, e.g., Vangsnes, 2001 on what he claims as the underlying NP structure for all languages based on the identification of functional categories). Another account, Lyons (1999), is described in chapter two.

To complete the account of articles as a group of grammatical markers of (in)definiteness, we have to address the paradoxical existence of zero. The following table describing the use of articles with count and noncount nouns is taken from Quirk et al. (1985:253), with modifications:

³ By reaching stage IV too, the English indefinite article is capable of expressing genericity, i.e. any unspecified member of a kind (cf. Carlson 1977). Stage V explains the co-occurrence of both articles in French.

Table 1.1: Use of articles with count and noncount nouns (alphabetical labels added)

		COUNT	NONCOUNT
SINGULAR	definite	(a) <i>the book</i>	(c) <i>the furniture</i>
	indefinite	(b) <i>a book</i>	(d) <i>furniture</i>
PLURAL	definite	(e) <i>the books</i>	
	indefinite	(f) <i>books</i>	

This paradigm of articles has been criticised by Dixon because number is insignificant for the characterisation of *the* (2006:34). In fact the table requires that mass nouns be interpreted in the singular. However, more importantly the table shows that an article is not used for indefinite mass nouns (d) and indefinite plural nouns (f). Can indefiniteness be signalled by absence of the definite article then? When we take into account singular proper names, this does not seem to be the case; consider ‘Tom chased Jerry’. The referents Tom and Jerry must be definite in this context although they are not determined by any article. One is then led to ask if no-article is neutral to definiteness.

When a morpheme expressing a certain meaning is grammaticalised, another related meaning (perhaps to signal the opposite) may come into existence by virtue of a similar morpheme being absent in a corresponding linguistic context (Bybee, 1994:235). This absence is thus a “meaningful zero” which is argued to have “positive semantic content,

which arises from the communicative context” (ibid.). A question that needs to be asked now is whether or not (d) and (f) above are marked by zero.

Bybee further postulates that a zero grammatical morpheme develops by the following mechanisms:

1. increase in frequency

Increase in use of zero correlates with increase in association of a given meaning and the absence of a grammatical morpheme;

2. inferential reasoning

This involves “the conventionalization of implicature” where an inference a hearer is allowed to make increasingly becomes part of the explicit meaning (1994:240).

It can be seen that frequency plays a large role in the grammaticalisation of zero. A related issue that needs further clarification is the origin of the meaning that comes to be expressed by a zero morpheme. According to Bybee, with no overt initiator or source available, “the semantic substance of zero [grammatical morphemes] must come from the discourse and cognitive context, which, of course, happens to be full of meaning,” and this is seen as being in line with the mechanism of inferential reasoning proposed above (1994:241).

Coming back to the cases of (d) and (f) in Quirk et al.’s paradigm, it can be said that the absence of an article in the referents *furniture* (mass) and *books* (plural) is a meaningful zero. In discourse, NPs refer to either the linguistic context or the situational context

(Quirk et al., 1985:253) or both (Fraurud, 1990) and have to be determined accordingly to indicate this status. It is proposed here that it was this obligatoriness that drove *a/an* to grammaticalise from *one* as a contrast marker to *the*. Assuming that this is correct, we see that in (8a) determination by *a* does not take place (and if it does, it will be effectively blocked) due to the mass conception of *furniture*. Similarly in (8b) it is constrained by the plural interpretation licensed by the morpheme *-s*.

8a. The shop sells *furniture*.

8b. The shop sells *books*.

Thus, it appears that zero corresponds to a meaningful absence of the definite article, rendering *furniture* as indefinite mass noun and *books* indefinite plural noun. The so-called zero article has now become a marker of indefiniteness.

Given the above account, how is the zero article distinguished from the absence of an article in proper names? A different perspective to this answer is discussed in 2.5.3 where Chesterman's theory of definiteness is discussed. However, it must be pointed out here that the use of a proper name as in 'She loves Tom' already indicates that the referent is definite. Thus there arises no need for determination by *the*. But when we focus on what can be conceived as a zero article in proper names, while it is fair to say that it corresponds to the absence of the definite article, it is not equivalent to a marker of indefiniteness. Having said this, if 'Tom' is on the other hand conceived as indefinite (i.e. not serving as a proper name), there will be two possibilities and the outcome is determined by the discourse context. First, in the absence of definiteness as a semantic constraint on the nominal, the singular indefinite article is called for, i.e. *a Tom*.

Second, because of the same absence of definiteness as a semantic constraint, the nominal has to be determined by *the* if it turns out that the discourse context actually requires definiteness marking, e.g. ‘the Tom (that she used to love)’.

The relevant computation above, over time, led to the grammaticalisation of the zero article as an indefiniteness marker for mass and plural. However the same cannot be said for the absence of articles in proper names. If this absence is to be conceived as a zero grammatical morpheme, it is *not* because the state of no-article is designed to mark indefiniteness. Nor is it used to mark definiteness because this work is already being done by the proper name. How do we escape from this impasse then? Perhaps the best way to characterise the absence of articles in proper names is not to analyse it in terms of articles at all, i.e. as it is done for person pronouns (e.g. I, you, he, she, they, etc.). It is for this reason that I agree with Chesterman’s term for it – null (although he still regards it as an article). As a conclusion to this section, the class of articles includes *the*, *a/an* and zero. The null article will be discussed again in 2.5.3.

1.3 Research questions

This study aims to address the following questions:

1. In naturally occurring data, what are the similarities and differences in the use of articles in Inner and Outer Circle varieties of English?
2. How can the differences be explained?
3. Where applicable, how do the findings of Outer Circle varieties compare to those of SLA?

1.4 Organisation of the chapters

This thesis is organised as follows: Following the present introductory chapter, the literature review on articles and definiteness is presented in chapter two. Chapter three reviews the process of nativisation in Outer Circle varieties. Chapter four describes the design of the study. The results are presented in chapters five, six and seven. Chapter five presents the findings of a frequency analysis of the articles. Chapter six discusses the collocational patterns of articles. Chapter seven examines the usage types of articles and highlights untypical usages found in some Outer Circle varieties. The thesis concludes in chapter eight.

Chapter 2 Articles and Definiteness

The purpose of this chapter is to describe the non-syntactic aspects of articles. These include their uses and their central meaning, i.e. definiteness. The term ‘definiteness’ is viewed either broadly or narrowly in the literature. ‘Definiteness’ in its broad sense is conveyed by determiners in general such as demonstratives, personal pronouns, and articles or even other linguistic means such as word order, stress, etc. In this thesis ‘definiteness’ is viewed in its narrow sense: it refers to definiteness as marked by special grammatical items such as articles (cf. simple definites and complex definites in Lyons, 1999).

The next section, 1.2, describes the uses of *the*, *a/an* and zero in the NPs that they determine. The meaning of ‘definiteness’ is discussed in section 1.3. It includes the four major definitions of definiteness as identified in the literature – uniqueness, familiarity, inclusiveness and identifiability, in that order. It will be shown that the architecture of each of these concepts is progressively dependent on the interrelationship between the speaker and the hearer, and the cognitive assumptions made between them. Other related notions of reference, specificity and genericity will also be discussed. Next, four proposals for a unifying account of the meaning of articles (i.e. Abbott, 2004; Huebner, 1983; Lyons, 1999 and Chesterman, 1991/2005) will be examined. The chapter concludes with a summary.

2. 1 Definite and indefinite articles

In chapter 1 it was proposed that the class of articles includes only *the*, *a/an* and zero.

In this section, we describe each of them in detail.

2.1.1 The definite article *the*

Quirk et al. propose that the use of the definite article *the* is “to mark the phrase it introduces as definite, i.e. as referring to something which can be identified uniquely in the contextual or general knowledge shared by speaker and hearer” (1985:265).

The description of the definite article *the* in terms of the unique identifiability of the referent by the interlocutors is found in other grammars as well. This position is most clearly articulated in Huddleston and Pullum (2002: 368-369). According to them, the definite article signals the speaker’s expectation that the learner is able to identify the referent. They claim that the kind of identifiability signalled by the definite article can be “understood in terms of pre-empting a question with *which?*” (p. 368). Thus in the sentence ‘I met the woman’ the hearer is assumed to be able to identify the woman without having to ask ‘Which woman?; if indeed the hearer responds by asking this question, the assumption has been wrong. Huddleston and Pullum (ibid.) also argue that identifiability as expressed by the definite article is “weak”. In their example ‘*The first person to run the mile in under four minutes was Roger Banister*’, they admit that it is difficult to pre-empt a *which* question for any of the referents in the sentence. However they still contend that the description can only be fulfilled by one entity by virtue of its

uniqueness (i.e. *the one person to have run the mile under four minutes for the first time*), and that it is in this sense that the referent is identifiable (ibid.).

For Huddleston and Pullum uniqueness is also a central notion in explaining definiteness because identifiability for count singular nouns is dependent on the fact that there is only one unique entity that satisfies the description of the referent. This can be seen in the examples given above. However they point out that there are also contexts where it is not necessary for only one entity to satisfy the description and strictly adhere to the uniqueness principle, e.g. ‘He married *the daughter* of his bank manager’ – uttered when the bank manager may have more than one daughter (p. 369).

Another dimension of meaning of the definite article as discussed by Huddleston and Pullum is existential presupposition. They point out that “use of the definite article presupposes the existence of the entity set, or quantity that the addressee is expected to be able to identify” (p. 369). All the examples cited above demonstrate this presupposition. However they add that, when used in a negative context, the definite article will still entail the existence of the entity, e.g., ‘He thinks it would be to his advantage to marry *the daughter* of his bank manager’ – uttered when the bank manager does not, in fact, have a daughter’ (ibid.).

For plural and non-count referents, they claim that the identifiability that is afforded by uniqueness can be achieved by the notion of totality. In the sentence ‘I have brought the boxes/*the syrup*’, uniqueness “applies now to a set or quantity” and its concern is “with

totality of the [boxes and syrup]” (ibid.). Also noteworthy is the fact that the totality expressed by the definite article is not as strong as that expressed by a universal quantifier. In ‘*All the bathroom tiles are cracked*’, it is understood that every tile is cracked, while in ‘*The bathroom tiles are cracked*’, the same meaning usually does not apply (p. 370).

Biber et al.’s (1999) and Downing and Locke’s (2002) definitions of definiteness, although differently phrased, similarly regard ‘uniqueness’ and ‘identifiability’ as the most important notions underlying its meaning. As observed by Quirk et al. (1985: 265) the (unique) identifiability⁴ of a given referent denoted by the definite article comes from shared contextual or general knowledge (‘endophoric’ or ‘exophoric’ in Downing and Locke’s functional terms, 2002). In Quirk et al. (1985: 265), the shared knowledge is further specified as follows:

- a. immediate situation e.g. *The roses* are very beautiful (said in a garden);
- b. larger situation (general knowledge) e.g. *the Prime Minister*;
- c. anaphoric reference-direct e.g. John bought a TV and a video recorder, but he returned *the video recorder*;
- d. anaphoric reference-indirect e.g. John bought a bicycle, but when he rode it one of *the wheels* came off;
- e. cataphoric reference e.g. *The girls sitting over* there are my cousins;

⁴ Lyons (1999: 15, fn 6) argues that ‘uniqueness’ and ‘identifiability’ are not one and the same. He claims that although many writers tend to use the term ‘unique identifiability’ to mean identifiability that is achieved through a certain degree of uniqueness, the two notions are distinctly separate. They have been developed separately and identifiability was first proposed as a counterargument to uniqueness. However, it is useful to note that combining these two notions is a practical thing to do (cf. Hawkins, 1978).

- f. sporadic reference e.g. My sister goes to *the theatre* every month;
- g. logical use e.g. When is *the first flight* to Chicago tomorrow?;
- h. use with reference to body parts e.g. Mary banged herself on *the forehead*.

The contexts in which the definite article is considered appropriate as described by Quirk et al. (1985) are generally similar to those in Huddleston and Pullum (2002) and Biber et al. (1999), although Biber et al. use fewer categories. Biber et al.'s corpus-based grammar (1999) adds another context in which the definite article is used, and this is one where referents are presented as if “familiar, though they have had no previous introduction” (p. 265), as can be seen below:

9. All this happened more or less. *The war parts* anyway, are pretty much true
<opening of a novel>

They claim that this particular use of *the* is a “way of quickly involving the reader in the story, inviting their cooperation in building up a mental picture of the fictional world” (ibid). Downing and Locke, using a functional perspective, explain that such use is occasionally found in literary discourse and it serves as an instance of Given information being exploited for special effects intended “to puzzle the reader and stimulate his interest in the narrative” (2002: 433).

The definite article is also found in fused head constructions (Huddleston and Pullum, 2002) e.g. ‘*the rich and the poor*’, in such uses as ‘He’s *the worse* for drink’ and ‘*the Atlantic*’ (Dixon, 2006: 37), as well as in many idioms, e.g. ‘kick *the bucket*’. A summary of all these uses is provided in table 2.1.

2.1.2 The indefinite article *a/an*

Singular count NP heads that are pronounced with vowel onsets require *an*. Otherwise, they occur with *a*. Quirk et al. (1985: 272) point out that the indefinite article is used “where the reference of X is not uniquely identifiable in the shared knowledge of speaker and hearer. Hence *a/an* is typically used when the referent has not been mentioned before, and is assumed to be unfamiliar to the speaker or hearer.” Below is their example:

10. *An intruder* has stolen *a vase*. The intruder stole the vase from a locked case. The case was smashed open (ibid).

Similarly for Downing and Locke (2002), the indefinite article is used “if there is nothing in the discourse or the situation or our general knowledge of the world which identifies [the entity] for us” (p. 429). According to Biber et al. (1999), the indefinite article “narrows down the reference of the following noun to a single member of a class and is often used to introduce a new specific entity in the discourse” (p. 260). Specificity, however, is not always a property of the indefinite article. This is noted by Biber et al., as well as by Downing and Locke (2002: 429-430). In Biber et al.’s example, ‘I’m looking for *a millionaire*, she says, but I don’t see many around’ (1999: 260), the speaker is not making any specific reference.

According to Quirk et al., a non-specific indefinite article is ‘non-referring’ in two ways. The first is found in the complement function of an NP in a clause. This is the non-referring use in which the article has a descriptive role (Biber et al.’s ‘classifying’), e.g. ‘Paganini was *a great violinist*’ (1985: 273). The second way in which an indefinite

article is non-referring can be found in an NP that does not appear to refer to “anything in reality at all”, e.g. ‘Leonard wants to marry *a princess who speaks five languages*’ (ibid). This example is similar to Biber et al.’s above. In relation to its non-referentiality, Huddleston and Pullum categorise the non-specific indefinite article as “non-quantitative” and note that the quantification, when available, is “generally non-proportional” (2002: 372) (cf. the exclusion of *some* from the group of articles in section chapter 1).

The quantitative meaning of the indefinite article can be traced to its origin – the unstressed form of *one*, e.g. ‘The Wrights have two daughters and *a son*’ (Quirk et al., 1985: 273). Quantitative *a* is especially recognisable in overtly quantifying expressions (e.g. ‘a hundred’, ‘a million’ and ‘a half’), non- numerical quantifiers (e.g. ‘a little’, ‘a large number of’) and measure phrases (e.g. ‘half *an hour*’, ‘ten dollars *a day*’) (Quirk et al., 1985: 274).

Additionally, Downing and Locke (2002: 430) note that the indefinite article may be used with proper nouns if there exist several entities bearing the same name, as can be seen in their examples below (cf. 1.2):

11. Is there *a John Smith* in this class? (indefinite; specific)
12. I was born on *a Monday* (indefinite; non-specific)
13. We had *a very hot June* last year (indefinite; specific)

They also comment that certain entities like continents, countries or cities may sometimes be “presented as indefinite, when something abstract or imaginary is predicated about it”⁵:

14. A politically united Europe will not be easy to achieve (ibid).

The indefinite article also occurs with a proper noun when the latter is used as a metonym:

15. I’ve got a Goya at home, but it’s only a reproduction (ibid).

In certain set expressions, *a* can occur with non-count nouns (Dixon, 2006: 42), as in:

16. He’s doing it with a vengeance.

All these uses are summarised in table 2.1.

2.1.3 The zero article

The zero article is regarded as a plural counterpart of the indefinite article (Quirk et al., 1985: 276; Biber et al., 1999: 261; Downing and Locke, 2002: 429). This can be seen in Quirk et al.’s example below:

17. Have you ever eaten roasted Ø chestnuts? (1985: 276)

When used as such, the number or amount of the referent is indefinite (Biber et al., 1999: 261). Biber et al. point out that the zero article is also used to classify plural nouns (e.g. ‘Two of his cousins are Ø teachers, his sister’s a teacher too’) (ibid). In fact, the zero article’s uses mostly express non-specific reference (cf. so-called ‘null article’ in 1.2).

⁵ However in ‘The rain forced the crowd to run for shelter, leaving *an embarrassed Jones* in his drenched suit at the podium’ the same use is extended to human beings.

Importantly, Biber et al. note that use of the zero article may indicate “neutralization of article distinctions” (ibid), which suggests why the phonologically null article is significant (Downing and Locke, 2002: 428-429). Quirk et al. (1985: 277) explain that the phenomenon is the result of the sporadic use (e.g. ‘She goes to *the theatre*’) having been “so institutionalised that the article is not used.” Biber et al. (1999: 261) further claim that nouns that are determined this way behave like count nouns in other contexts. Biber et al.’s (1999: 261-263), list of uses of the zero article in this manner are given below⁶:

- a. meals, e.g., Are they going out for Ø dinner or something? (cf. Bye, bye dear, thanks for *the lunch* – the definite article is used when a special meal is singled out);
- b. institutions, e.g., They’re in Ø hospital, badly injured (BrE) (cf. It’s the diamond jubilee of the hospital);
- c. predicatives with unique reference, e.g., Lukman was re-elected OPEC Ø president in November;
- d. means of transport and communication, specifically with the use of ‘by’ in a PP, e.g., go by Ø bus (cf. He saw me off on *the bus*);
- e. times of the day, e.g., Tomorrow at Ø dawn we’ll begin our journey (cf. She sat and waited for the dawn);
- f. days, months, and seasons, e.g., It was on the radio on Ø Sunday (cf. That was *the Sunday* before we moved);
- g. parallel structures, e.g., He travelled from Ø country to Ø country.

⁶ In Downing and Locke’s view, however, many of these uses refer to the whole class, and are thus generic (e.g. It’s time for *bed*; He’s in *hospital*).

- h. block language, e.g., *Ø Fire* kills *Ø teenager* after *hoax* (cf. *A teenager* died in *a blaze* at his home after firemen were diverted by *a call* that turned out to be *a student prank*);
- i. vocatives, e.g., See you later, *Ø mate*.

These uses are summarised in table 2.1.

2.1.4 The generic articles

Generic reference is concerned with a whole class (Quirk et al., 1985; Biber et al., 1999; Downing and Locke, 2002). All three forms of article may be used generically.

Used in this sense, the indefinite article selects any representative member of the class:

18. The best way to learn *a language* is to live among its speakers (Quirk et al.; 1985: 281).

It follows that the indefinite article cannot be used attributively in describing the class as a whole, e.g. **A tiger* is becoming extinct' (ibid).

By contrast, when used generically, the zero article identifies the whole class:

19. *Ø Cigarettes* are bad for your health (Quirk et al., 1985: 282).

For Downing and Locke, however, the zero article may imply “*all or most* members of the class of entity” (2002: 432). Biber et al. (1999: 265) point out that the use of the zero article with non-count nouns can result in reference that remains generic even when there exists premodification:

20. In Ghana *Ø coconut wine* is not as popular and common as palm wine (ibid).

The definite article, meanwhile, has the most limited ability to function generically (Quirk et al.; 1985: 282). When used (usually in literary discourse or formal style), it identifies “the class as represented by its typical specimen”:

21. A great deal of illness originates in *the mind* (ibid).

Downing and Locke, however, observe that “*the* tends to generalise more readily than *a/an*” and that “the [use of *the* and singular count noun] may have a generalising value, even when not used in a generic statement” (2002: 434). The examples below serve to illustrate this:⁷

22. Do you play *the piano*?

23. Some people sit for hours in front of *the television*.

Generic *the* is also used with nouns denoting nationality, e.g. ‘the Chinese’ (Quirk et al., 1985: 284; Biber et al., 1999: 266), and in phrases where the adjective is also the head that describes a group of people e.g. ‘the rich’ (Quirk et al., 1985: 284; Downing and Locke, 2002: 434) (cf. fused-head constructions in 2.1.1). However, Downing and Locke warn that “not all adjectives can function in these ways and the non-native speaker should be cautious in choosing them” (ibid).

It is also noted by Downing and Locke that the use of generic *the* and *a/an* can be strengthened by the presence of a modifier/qualifier or “a clausal Adjunct such as *usually*” (ibid)⁸:

⁷ Presumably, the sentences are said in a context where the referent (i.e. piano or television) is physically present.

⁸ It should be pointed out, however, that (25) and (27) are less likely than (24) and (26) respectively.

24. An unhappy child usually cries a lot.
25. The unhappy child usually cries a lot.
26. A driver who drinks is sometimes dangerous.
27. The driver who drinks is sometimes dangerous.

Overall, however, they regard the use of the zero article and plural nouns as representing the most common way of expressing genericity (2002: 433). Hence, these examples:

28. Kangaroos are common in Australia.
29. Unhappy children cry a lot.
30. Drivers who drink a lot are dangerous.

The uses of generic articles are summarised in table 2.1.

Table 2.1: The uses of *the*, *a/an*, zero and generic articles

	referential	non-referential	generic
the	<ul style="list-style-type: none"> • refers to something that can be identified uniquely in the contextual or general knowledge shared by speaker and hearer, specified as follows: <ul style="list-style-type: none"> ○ immediate situation e.g. <i>The kids</i> are quiet (said in a classroom) ○ larger situation (general knowledge) e.g. <i>the Prime Minister</i> ○ anaphoric reference – direct e.g. I brought a book and a pencil but broke <i>the pencil</i> later ○ anaphoric reference – indirect e.g. I brought a book but <i>the pages</i> were missing ○ cataphoric reference e.g. <i>The band</i> on stage is Stereophonics ○ sporadic reference e.g. I take <i>the train</i> to work everyday ○ logical use e.g. <i>The last person</i> to leave will lock the door ○ reference to body parts e.g. He hit me on <i>the head</i> • presupposes uniqueness (a count noun typically has only one unique entity to identify) although cf. ‘He married <i>the daughter</i> of the butler’ (even though the butler may have more than one daughter) • presupposes existence of the entity of the referent to be identified; in a negative context <i>the</i> still entails existence, e.g. ‘He thinks it’s good to marry <i>the daughter</i> of the butler (although t the butler doesn’t have a daughter) • unique identifiability is achieved via totality for plural and non-count nouns; this totality is not as strong as a universal quantifier, e.g. <i>all</i> • refers to something presented as if familiar although without previous introduction, e.g. ‘Jalal sat on <i>the bus</i>, staring at <i>the empty house</i> and the life he was about to leave behind’ 	<ul style="list-style-type: none"> • occurs in idioms, e.g. ‘grab <i>the bull</i> by its horns’ 	<ul style="list-style-type: none"> • refers to a class as represented by its typical specimen, e.g. ‘<i>The monkey</i> is a curious animal’ • is used with nouns denoting nationality, e.g. <i>the Finns</i> • occurs with the fused-head construction of adjective and noun, e.g. <i>the rich</i> and <i>the poor</i>

a/an	<ul style="list-style-type: none"> • refers to something that is not uniquely identifiable in the shared knowledge of speaker and hearer; introduces a new specific entity into discourse e.g. 'A <i>rock</i> dropped from the sky today' • carries the quantitative meaning of 'one', e.g. 'I have <i>a dog</i> and <i>a cat</i>' • is used with a proper noun to individualise each referent if there exist several bearing the same name; specific use, e.g. 'Is there <i>a Mary Jackson</i> here?' • is used with entities like continents, countries or cities to create an indefinite referent with an abstract or imagined quality, e.g. '<i>A stronger Asia</i> will emerge from this crisis'. And also with humans, e.g. 'They left, leaving <i>a puzzled Christine</i> at the door' • occurs with a proper name to be used as a metonym, e.g. 'There was <i>a Carolina Herrera</i> among her dresses in the wardrobe' 	<ul style="list-style-type: none"> • creates non-specific reference, e.g. 'He's looking for <i>a wife</i>' • carries a descriptive role or classifies, usually non-quantitatively, e.g. 'May is <i>a lousy singer</i>' • is used with a proper noun to individualise each referent if there exist several bearing the same name; non-specific use, e.g. 'She'll be born on <i>a Friday</i> next year' • is used with non-count nouns in certain set expressions, e.g. 'The rain fell with <i>a vengeance</i>' 	<ul style="list-style-type: none"> • refers to any representative member of the class; cannot be used attributively to describe a whole class, e.g. '<i>A camel</i> can go without water for months' but '*<i>A camel</i> is becoming extinct'
------	--	---	---

ZERO	<ul style="list-style-type: none"> • indicates neutralisation of article distinctions; sporadic use of <i>the</i> is so institutionalised that the article is dropped, as in: <ul style="list-style-type: none"> ○ meals e.g. What's for <i>dinner</i>? ○ institutions e.g. She goes to <i>church</i> sometimes ○ means of transport and communication e.g. I'm coming by <i>bus</i>. ○ times of the day e.g. At <i>night</i> they roam the earth ○ days, months and seasons e.g. Come and see me on <i>Monday</i> ○ accompanies a unique role or task e.g. Claire Jones is <i>Professor of Nuclear Science</i> at Cambridge University. ○ parallel structures e.g. He went from <i>door to door</i>, showing them his wares ○ block language e.g. <i>Singer</i> arrested for sex with <i>minor</i> ○ vocatives e.g. You look great, <i>darling</i>. 	<ul style="list-style-type: none"> • used as a plural counterpart of the indefinite article; the number and amount of the referent is indefinite, e.g. 'There are <i>persimmons</i> in that box' • classifies plural nouns, e.g. 'We are <i>singers</i>' 	<ul style="list-style-type: none"> • refers to the whole class, e.g. '<i>Rainy days</i> make you sad'
------	---	--	--

2.2 What is the meaning of definiteness?

The uses of articles above arise from their status as overt markers of definiteness. However it is useful to remember that some languages do not encode definiteness. This fact leads us to a very complex question: If some languages do not even mark definiteness, what does it mean for an NP to be definite or indefinite in English?

Consider the following sentences:

31. Adam talked to a girl.

32. Adam talked to the girl.

What is the difference in meaning between (31) and (32)? A common answer would be that the girl in (32) is not just any girl – she is somehow more ‘known’ than the first one. This answer is expected, given one would be guided by the term ‘definite’ and the presence of the definite article *the* in (32). However, the girl in (31) may not always be just any girl – some girl that was unheard of prior to the utterance. In a context where the speaker has a particular girl in mind (e.g. a girl that he is acquainted with), sentence (31) would be just as felicitous (cf. specificity in 2.3). So the use of the indefinite article in (31) appears to belie the meaning of its name.

Various concepts have been proposed to capture the notion of ‘definiteness’. The following sections will show that any attempt to reduce this category to one unitary concept is usually not adequate. Below I will describe the main definitions that have been put forward in the previous approaches to the study of definiteness: uniqueness, familiarity, inclusiveness and identifiability. Our own working definition will be based on certain elements from each of them.

2.2.1 Uniqueness

In the following sentences, one can easily understand why uniqueness has been proposed as the definition of definiteness as encoded by the definite article:

33. The Prime Minister of Malaysia is the son of a previous prime minister.

34. ?A Prime Minister of Malaysia is the son of a previous prime minister.

In uttering (33), the speaker has in mind the one person who is constitutionally the prime minister of Malaysia, and at the time of writing this refers to Najib Razak. Upon hearing (33), the hearer knows which person is meant by the speaker by virtue of the uniqueness of the reference. In contrast, in uttering (34) there is an implication that there exists at least another prime minister in the country, hence the questionable acceptability of the sentence. Another example that might convince us that uniqueness is indeed definiteness is given by Abbott (1999: 2): ‘That wasn’t *a reason* I left Pittsburgh, it was *the reason*’. The use of *the* in *the reason*, as a contrast to *a reason*, reinforces the meaning that there exists only one such reason.

The idea that uniqueness is the meaning of definiteness can be traced back to the philosopher Russell (1905) who argued:

The only thing that distinguishes “the so-and-so” from “a so-and-so” is the implication of uniqueness. We cannot speak of “the inhabitant of London,” because inhabiting London is an attribute which is not unique. We cannot speak about “the present King of France,” because there is none; but we can speak about “the present King of England.” Thus propositions about “the so-and-so” always imply the corresponding propositions about “a so-and-so,” with the addendum that there is not more than one so-and-so (p.176).

Russell's theory of definiteness was concerned with the logic of sentences containing definite descriptions, or definite NPs.⁹ Although Russell's analysis seemed to be able to explain many uses of the definite article, it has some problems.

One problem with Russell's theory is its claim that, on the use of *the*, the existence of the entity is asserted (i.e. that there exists a prime minister and this prime minister is unique in 33). Strawson (1950) pointed out that, on the use of the definite article, such existence is not asserted; it is merely presupposed. Using (33) as an illustration, what is presupposed is that there is a Prime Minister of Malaysia and there is no more than one person who is a Prime Minister of Malaysia. What is asserted is that he is the son of a former prime minister. Strawson's claim is based on the observation that, given a sentence like 'The king of France is bald' (Russell's famous example) in a context where it is known for a fact that France no longer has a monarchy,¹⁰ what must be negated is not the existence (i.e. saying 'That's false' and referring to the non-existence of this king), but the assertion that he is bald (i.e. saying 'The king of France is not bald'). If someone were to say this sentence out loud in a public place, a Good Samaritan would probably respond by only pointing out the hapless speaker's confusion in thinking that the king actually existed. If the same speaker were to say 'The Prime Minister of Malaysia is a go-go dancer', many would readily respond 'No,

⁹ Formally, an indefinite NP such as (i) is analysed as (ia) while a definite NP such as (ii) is analysed as (iia). (ib) and (iib) are respectively their verbal descriptions:

- i. A man came.
- ia. $\exists x[\text{Man}(x) \ \& \ \text{Came}(x)]$
- ib. There exists something which is both a man and came.
- ii. The man came.
- iia. $\exists x[\text{Man}(x) \ \& \ \forall y[\text{Man}(y) \ \rightarrow \ y=x] \ \& \ \text{Came}(x)]$
- iib. There exists one and no more than one thing which is a man, and that thing came.

¹⁰ Russell's aim was to demonstrate that although this sentence is meaningful, it is false.

he is not!' In terms of Russell's original concern then (see footnote 10), the sentence 'The king of France is bald' is neither true nor false; it is just meaningless.

Strawson's criticism of Russell can also be taken to mean that a definite NP is referential (as opposed to Russell's 'quantificational') because the use of 'The king of France' must refer to an entity (hence, the Good Samaritan's response). This is also the view of Donnellan (1966). However, in his own criticism of Russell, Donnellan points out that Strawson was mistaken in assuming that a definite NP is always referring. In his analysis, Donnellan shows that a definite NP can be either referring or 'attributive' (non-referring) depending on the context. In his example, the question 'Who is the man drinking a martini?' (even though the man is actually drinking water in a martini glass) can be answered (e.g. 'Oh, he's my long-lost cousin and he's drinking water, actually') because it is intended referentially and picks out a particular person. If used attributively, this question cannot be answered because the man being talked about is not unique by virtue of not drinking a martini. However it is also important to note that, where the context permits, the same question can be used attributively to find one such person in say, a party where everyone else is drinking water.

The third criticism of Russell's uniqueness theory is the problem of incomplete descriptions. This refers to referents of NPs that are not unique on a universal scale. In the everyday world, such referents are many: 'the glass (on the table)', 'the rug (on the floor)', 'the man (at the door)', and so on. Given Russell's formulation that the referent of a definite NP must be unique to the extent that there must not exist other entities in

the world fitting the descriptive content of the NP, the glass, table, rug, floor, man and door above, and countless other entities in this world cannot be considered unique in a true sense (Hawkins, 1978: 96). This presents a serious problem to Russell's formulation of uniqueness-as-definiteness theory.¹¹

There have been many attempts at a solution (see Abbott, 2004 for a review). One of the most well known is Hawkins (1991) in which he describes a pragmatically reduced context known as 'Pragmatic sets' or 'P-sets'. This is to enable a more realistic context in which the referents are to be evaluated for their uniqueness. However, using McCawley's example (1979 in Lewis, 1979), Lewis points out that even delineating such a context requires attention to even further detail such as 'salience':

It is not true that a definite description 'the F' denotes x if and only if x is the one and only F in existence. Neither is it true that 'the F' denotes x if and only if x is the one and only F in some contextually determined domain of discourse. For consider this sentence: ... 'The dog got in a fight with another dog' (McCawley). ... [For it] to be true... 'the dog' must denote one of two...dogs, both of which belong to the domain of discourse. The proper treatment of descriptions must be more like this: 'the F' denotes x if and only if x is the most salient F in the domain of the discourse, according to some contextually determined salience ranking.
(Lewis, 1979: 348)

The problem of incomplete descriptions continues to be debated by linguists, although it has to be said that in each proposal for a solution, a significant amount of attention is

¹¹ In defence of Russell's theory, Neale (1990) highlights the fact that the problem of incomplete descriptions is not limited to definite NPs but it affects all quantified NPs. He gives the example of *Everyone was sick*, uttered as a response to a query about a party he went to the previous night. He elaborates:

Clearly I do not mean to be asserting that everyone in existence was sick, just that everyone *at the dinner party I had last night* was. [...] Similar examples can be constructed using 'no', 'most', 'just one', 'exactly eight', and, of course, 'the'. [...] Indeed, the problem of incompleteness has nothing to do with the use of definite descriptions *per se*; it is quite a general fact about the use of quantifiers in natural language (p. 950).

given to the contextual aspect of definiteness (Bach, 1999), which lies in the domain of pragmatics.

Incomplete descriptions do appear to present the most problems for the uniqueness theory of definiteness. Other than incomplete descriptions, there are other uses of the definite article that do not seem to fulfil the uniqueness requirement of the theory. Du Bois (1980: 233), for example, gives the following sentence:

35a. The boy scribbled on *the living room wall*.

Why is the use of *the* acceptable even though out of the typical set of four, the one wall that the boy scribbled on is not singled out for definiteness? In fact, as pointed out by Du Bois (ibid.), any attempt at ‘correcting’ this sufficiency will result in oddness:

35b. ?He scribbled on *a living room wall*.

35c. ?He scribbled on *the north living room wall*.

There are other examples that display similar discrepancies:¹²

36. Towards evening we came to *the bank of the river*. (From Christophersen, 1939: 140)

37. [Hotel concierge to guest, in a lobby with four elevators] You’re in Room 611. Take *the elevator* to the sixth floor and turn left. (From Birner and Ward, 1994: 93)

¹²Lyons (1999: 14) gives a few examples where he claims inclusiveness (a term due to Hawkins, 1978 as an extension of uniqueness) does not work. Here, I will discuss one:

[Two academics] A: How did the seminar go?
 B: Fine. *The student* gave an excellent presentation, which generated a really good discussion, with all the other students contributing well.

However it seems that the definite NP *the student* here does not run into any problem with the uniqueness requirement. Of the set of students the two academics are talking about, only one student gave a presentation and an excellent one at that. This is the student singled out for definiteness by *the*.

These NPs display non-uniqueness that they can be said to appear as indefinite. Carlson and Sussman (2005) furnish psycholinguistic evidence to prove that this is indeed the case. Their experiment shows that subjects did indeed take a longer time to mentally process NPs such as the above. However, in agreement with Birner and Ward (1994), Abbott (2001: 10-11) argues that the uniqueness of the referents in (35a) through to (37) can be explained in terms of location. Abbott's rebuttal will be elaborated in 2.4.1 below where her proposal for uniqueness as a unifying account of definiteness will be presented in detail. (Du Bois' own explanation will be presented in the discussion on 'identifiability' in 2.2.4.)

Another problem with Russell's uniqueness theory is its neglect of plural and mass nouns. This shortcoming has been addressed by Hawkins (1978; 1991) in his conception of 'inclusiveness'. This will be elaborated in 2.2.3 below.

In summary, the uniqueness theory has significantly furthered our understanding of definiteness and the use of articles as its markers. However, this is not a theory that is universally accepted. Other than the problems highlighted above, the uniqueness theory is in competition with a different theory that approaches definiteness from a very different perspective. This is the familiarity theory.

2. 2. 2 Familiarity

The familiarity theory of definiteness was originally proposed by the Danish linguist Christophersen (1939) who asserts that the notion of definiteness must be associated

with some kind of previous knowledge. Contrary to Russell's uniqueness, Christophersen's formulation of familiarity frames definiteness in explicit hearer-based terms. He elaborates:

Now the speaker must always be supposed to know which individual he is thinking of; the interesting thing is that the *the*-form supposes that the hearer knows it too. For the proper use of the form it is necessary that it should call up in the hearer's mind the image of the exact individual that the speaker is thinking of. If it does not do that, the form will not be understood (Christophersen, 1939: 28).

This approach does explain a great number of uses of the definite article. Recall (32) where the nominal *girl* can be supposed to be familiar to the hearer of the utterance. Familiarity with this girl on both the speaker's and hearer's parts allows the felicitous use of *the*. Notice also that familiarity, or the lack thereof, also explains why *a* is used in (31). The familiarity theory is also very effective for explaining definite NPs that have been introduced before:

38. I live next to a scientist. *The scientist* keeps to himself though.

Importantly, in (38) it can be seen that familiarity can be established on the basis of the given text. A particular achievement of this theory is that it also explains many referents in definite NPs that are established on a non-textual basis.¹³ This is exemplified by *the author* in 'The book is so ridiculous - the author must be crazy'. It is possible to mark *author* as definite because our familiarity with books tells us that each of them typically has an author. This is what Christophersen calls the "implicit contextual basis" (1939: 29). In addition, there is the "situational basis" (1939: 30),

¹³ This kind of NPs has attracted the attention of many linguists, e.g., Lewis (1979) who discusses them under 'accommodation', Clark and Haviland (1977) 'bridging inferences' and Prince (1981) 'inferables'.

where familiarity is based on a larger, extended context of the utterance. For instance, upon mounting a bus, one can talk of the driver, the passengers, the seats and so on.

In formulating familiarity as the definition of definiteness, Christophersen has brought in many pragmatic properties to bear on the semantics of articles. Despite such a novel attempt to elucidate the notion of definiteness, it has been repeatedly shown that familiarity is not a sufficient condition to explain the meaning of *the*. Some uses of *the* show that the definite article is frequently used to mark ‘unfamiliar’ referents. Below are two examples:

39. If you’re going into the bedroom, would you mind bringing back *the big bag of potato chips that I left on the bed?* (From Birner and Ward, 1994: 93)

40. What’s wrong with Bill? Oh, *the woman he went out last night* was nasty to him.
(From Hawkins, 1978: 101)

Mention of the big bag of potato chips in (39) and the woman in (40) cannot be said to be based on familiarity.¹⁴ There is no previous knowledge of these entities for the hearer to rely on.

Birner and Ward (1994: 94) also point out that in cases such as *the author* above, familiarity does not ultimately explain the felicitous use of *the*. According to them

¹⁴ If familiarity of *bag of potato chips* and *woman* is assumed by only the speaker and the hearer is expected to accept this condition, then one could practically determine all referents, definite or otherwise, in this way: Have you seen *the boys?*, Please give me *the pencil*, I’m running towards *the tree*, when none of these referents is familiar to the hearer. Since this is not the case, proponents of uniqueness would argue that (39) and (40) support their theory. Notice also the cataphoric information (e.g. ...he went out last night) in (40) does not amount to familiarity. As a definition of definiteness, familiarity cannot explain above and beyond what the word means. It seems that proponents of either uniqueness or familiarity tended to argue for their theory to the total exclusion of the other. In hindsight this practice did not appear to be beneficial.

given that a book typically has one author, one can utter ‘The book is ridiculous; the author must be crazy’ and safely assume that the hearer knows this fact. (In short they claim it is uniqueness, not familiarity that sanctions the definite article.) Christophersen himself acknowledges this problem in his own way: “Now, in all strictness, this term (familiarity) is not always quite correct. Though the previously acquired knowledge may relate to the very individual meant, yet it is often indirectly that one is familiar with what is denoted by the word” (1939: 72-73). Birner and Ward provide a further example to prove that not only is familiarity not a sufficient condition for the felicitous use of the definite article, but also that it is not necessary:

41. Professors Smith and Jones are rivals in the English Department, and each of them has received a major grant for next year. ?The other members of the department are very excited about *the grant*. (1994: 94)

It can be seen there that although a mention of *grant* has been made prior to the definite *the grant*, the supposed familiarity does not work in favour of the definite article. In fact, they claim that the presence of two grants, one for each professor, only results in confusion that prevents the hearer from knowing which one is being referred to.

Familiarity is also found to be problematic in explaining non-referential NPs, which refer to NPs that do not “pick out a referent” but are predicational in the sense that they “denote a quality or characteristic” (Declerck, 1986:28). This is illustrated below:

42. Chan is *a scientist*.

43. Chan is *the leader*.

44. Chan is *the acme of courtesy*. (Adapted from Declerck, 1986).

In the above sentences, familiarity cannot account for neither the definiteness of the italicised NPs in (43) and (44) and nor can it account for the indefiniteness of that in (42).

The problems listed above present a serious challenge to the familiarity theory of definiteness. However, many linguists agree that the context of utterance is an important criterion in any effort to understand what definiteness is. These linguists have worked on this theory further and many of them have clearly articulated their improved position through a notion that they call 'identifiability'. This is to be described in 2. 2. 4 below.

2.2.3 Inclusiveness

Hawkins' (1978) work on definiteness is noted for his notion of 'inclusiveness'. It is one that extends, quite naturally, from uniqueness. While Russell's conception of uniqueness is based on singular count nouns, inclusiveness works to cover plural and mass nouns. Consider Hawkins' examples below (1978:159):

45. Bring *the wickets* in after the game of cricket.

46. I must ask you to remove *the sand* from my gateway.

In each sentence, the referent of the italicised NP is unique, but only in reference to the whole set. This means if the hearer in (45) were to bring in only four (from the total set of six) wickets, he would be assumed to misunderstand the denotation of *the wickets*. Similarly, if the hearer in (46) were to remove only some of the sand, the same degree

of miscommunication would occur. In the case of ‘the sun’, ‘the boy’, and other singular NPs, their inclusiveness is restricted to the one member that constitutes the set.

The theory that Hawkins introduces the notion of inclusiveness is called ‘location theory’. Its mechanism is similar to that of identifiability, which will be explained next.

It must be pointed out, too, that Hawkins does not simply base this identifiability on the mere uniqueness of a referent:

According to location theory the speaker performs the following acts when using a definite article. He (a) introduces a referent (or referents) to the hearer; and (b) instructs the hearer to locate the referent in some shared set of objects...; and he (c) refers to the totality of the objects or mass within this set which satisfy the referring expression (1978:167).

Hawkins’ ‘shared set of objects’ is drawn upon the notion of ‘previous knowledge’, which is not unlike that of Christophersen.

2.2.4 Identifiability

As noted in 2.2.2, identifiability as the definition of definiteness came about as a further development of the familiarity theory. Basically the use of the definite article enables the hearer to identify the referent intended by the speaker. Before this concept is described in detail, a potentially puzzling treatment of identifiability that is found in the literature must be highlighted. This is the term ‘unique identifiability.’

‘Unique identifiability’ is used to mean that a referent is identified through its unique being in the context of utterance. This term is found (and variously defined to this effect) in, among others, Leech (1983), Quirk et al. (1985), Gundel et al. (1993), Birner

and Ward (1994) and Epstein (2002). Although intuitively appealing, it has been pointed out by Lyons (1999:15, fn. 6) that the term suggests, for these writers at least, that the uniqueness theory and the identifiability theory are one and the same, obscuring the fact that identifiability is originally developed from familiarity. This can be problematic considering that in many other analyses the two theories of uniqueness and familiarity are necessarily differentiated and the opposition between them is often discussed in detail, upon which many further insights are drawn (cf. Hawkins' 'location theory' in 2.2.3).

For the purposes of this study, the term 'identifiability' that is reviewed in this section refers to the familiarity-related notion that was introduced in 2.2.2. As noted earlier, familiarity is found to be too limited to apply in, for example, (39) or (40). Lyons (1999:3) provides similar cases. One example is given here:

47. They've just got in from New York. *The plane* was five hours late.

Lyons argues that in the context of this sentence, it is interesting that *the plane* is sanctioned by something other than an NP (i.e. the phrase *got in from New York*). It is also interesting to note that travelling from New York to Manchester (Lyons' example) does not necessarily involve aeroplanes. One could, for example, take a cruise ship. In the face of these facts, how can one say that *the plane* is somehow familiar? That one can say so requires much qualification in the original argument. This is: presently, the most common way to travel from New York to Manchester is by flying. Thus, familiarity in its strict sense is limited to explain the use of the definite article. But using an expanded form of previous knowledge (linguistic or non-linguistic), one is

able to identify the plane. It may necessitate going beyond a one-to-one association between a referent and its recognition (in the loose sense of the word). This is what is meant by identifiability. One could argue that uniqueness also plays a part in identifiability (hence, unique identifiability) but Abbott emphasises that “use of the definite conveys to the addressee that they ought to be able to determine a unique referent from the description used *plus contextual or background information, whether or not they had prior acquaintance with it*” (2004:13, emphasis added). Birner and Ward maintain that:

What is required for felicitous use of the definite article (and most uses of other definites) is that the speaker must believe that the hearer is able to *individuate* the referent in question from all others within the discourse model (1998:122, emphasis in original)

Notice also that *the* by itself does not identify, but “invites the hearer to exploit clues in the ... context to establish the identity of the referent” (Lyons 1999: 6, fn. 3).

Identifiability is effective in explaining the use of the definite article on what Christophersen calls ‘contextual’ or ‘situational’ bases. However, recall sentences (35) – (37) which are problematic for familiarity; it appears that they are equally problematic for identifiability. In an effort to solve this problem, Du Bois (1980) formulates what he calls the “curiosity principle”. This principle states that “a reference is counted as identifiable if it identifies an object close enough to satisfy the curiosity of the hearer” (1980: 233). Using (35a) as an example, this principle explains why it is acceptable for the speaker not to give full identification to the wall that was scribbled

on by the boy. In fact, we get by without full identification of many referents that we use in our everyday speech (Du Bois 1980: 233).

Before we move on to the next section, it is useful to note that uniqueness has been extended to inclusiveness, and identifiability can be seen to subsume familiarity. Lyons proposes that identifiability suits definite referential NPs better while inclusiveness (uniqueness) more properly captures definite non-referential NPs (1999: 253). It can be seen that uniqueness (in Russellian terms) is semantic in nature while the idea behind familiarity is couched in discourse-pragmatic terms (Abbott, 2004:12).

2.3 Reference and specificity

Reference is “the relationship which holds between an expression and what that expression stands for on particular occasions of its utterance” (Lyons, 1977: 174). It used to be thought that all words (or any linguistic expressions) have referents (or extensions) in the real world. However, this view is simply problematic because language allows the use of obviously non-referential words like ‘unicorn’, ‘utopia’, ‘Armageddon’, etc. in the real world. A more commonly held view now is that the referents of linguistic expressions are more abstract in nature, often glossed as ‘mental entities’ and captured in similarly abstract concepts of “universe of discourse” (Givon, 1984) or “mental spaces” (Fauconnier, 1985; 1998).

A concept closely related to reference is specificity. It is concerned with whether or not the reference of an NP involves a specific entity. To illustrate, recall the girl in (31), which is reproduced below:

31. Adam talked to *a girl*.

It was pointed out above that this girl may or may not be a particular girl that the speaker has in mind, with no assumption of hearer-identifiability. If this girl is a particular girl in the mind of the speaker, then the referent intended here is specific. As summarised by Ionin et al., specificity is “speaker intent to refer” (2004: 5). While English does not have special markers for specificity, *this* in informal English is capable of expressing this meaning, e.g. ‘Adam talked to *this girl*, but he wasn’t interested in her at all’.

Reference by definite and indefinite NPs, however, is a contentious issue in the literature. Some linguists believe that definite NPs refer and that indefinite NPs do not. In the case of indefinite NPs, it is believed that they only describe (Donellan, 1966; Fodor and Sag, 1982). Other linguists believe that both definite and indefinite NPs are non-referring in the semantic sense (Kripke, 1977; Neale, 1990 and Ludlow and Neale, 1991). For these linguists, reference, in the sense of picking out a referent, is carried out only by proper nouns, demonstratives and personal pronouns. On this view definite and indefinite NPs only describe, “[denoting] whatever meets the description (definite) or something which meets the description (indefinite)” (Lyons, 1999: 166). Following Russell (1905), this view considers definite and indefinite NPs ‘quantificational’. While these NPs do have referential uses, it remains that they themselves do not refer: “Thus

while the proposition representing the literal meaning of a sentence containing a definite or indefinite description will involve quantification, this sentence may be used to convey a proposition involving direct reference” (Lyons, 1999:166). What this means is that reference is a matter of pragmatics rather than semantics (ibid). In view of the lack of agreement, Lyons does not commit himself to any of the strong views. For him, definite and indefinite NPs are potentially referring but need not do so (cf. Epstein 1992; 2002, on his argument that restricting the study of articles to their referential uses is ignoring some other functions that can be fulfilled by them).

In the case of specificity, a common observation that can be made in regard to indefinite NPs is that they can be either specific or non-specific (see sentence (31)). However, Lyons (1999:168) points out that the same can be observed for definite NPs. Consider his examples:¹⁵

48 a. I’m going to have lunch with the president tomorrow – I’m dreading it, he’s such a boring man.

48 b. I’m going to have lunch with the president tomorrow – that is, if the election takes place today and we have a president.

49 a. We can’t start the seminar, because the student who’s giving the presentation is absent – typical of Bill, he’s so unreliable.

¹⁵ Lyons’ examples are intended to contrast readings in opaque (i.e. 48) and non-opaque (i.e. 49) contexts. The point he wants to make is that the presence of a logical operator in (48b) and its absence in (49b) explain the different labels he uses to describe the specificity/non-specificity in these two contexts: wide-scope for (48a), narrow-scope (48b), referential (49a) and non-referential (49b). Note that for each context, an indefinite NP can also be specific or non-specific.

49 b. We can't start the seminar, because the student who's giving the presentation is absent – I'd go and find whoever it is, but no-one can remember, and half the class is absent.

A possible way to view this phenomenon, as observed by Lyons, is to regard specificity as an ever-present ambiguity in the context of utterance. It is the speaker, when s/he has a particular referent in mind, who determines whether to indicate this knowledge explicitly to the hearer or not. Similarly Ionin et al. claim that “the specificity distinction is independent of the definiteness distinction” (2004: 9). Thus we need to rely heavily on pragmatics to resolve the arising ambiguities.¹⁶

2.4 Genericity¹⁷

A different kind of use of articles is to express genericity (see 1.3.4). A generic sentence usually “contains a kind-referring noun phrase as its topic and a characterizing predicate which expresses a prototypical (but not necessarily essential) property of the topic” (Behrens, 2005:275). It is crucial to bear in mind that languages do not encode genericity in a uniform manner. This has important consequences on the kinds of generalisations we can make based on the patterning of genericity observed in English alone (Behrens, 2005).

In English, genericity can be expressed using any of the articles. This is illustrated below:

¹⁶ There are languages whose articles have grammaticalised into markers of specificity, e.g., Samoan (Lyons, 1999). Informal English has one such marker too, i.e. *this*.

¹⁷ The term ‘genericness’ is also found in the literature. I will use ‘genericity’ throughout this thesis due to its more frequent use.

50. The coconut tree is very useful.

51. Coconut trees are very useful.

52. A coconut tree is very useful.

However, notice that the definite plural is not usually used for expressing genericity, e.g., ‘*The coconut trees* are very useful’. According to Lyons (1999:181-182), as a marker of genericity, the definite plural can be found with NPs headed by “nationality nouns and some nouns denoting classes of classes”, e.g., ‘*The Brits* don’t eat much rice, do they?’ Or ‘*The dinosaurs* were probably the first creatures on Earth’ (see table 2.1).

Hawkins argues that the indefinite article “refers exclusively” even in generic use (1978: 214). Thus, use of this article implies “any representative member of the class” (Quirk et al. 1985: 281). As mentioned earlier in the chapter, the indefinite article cannot be used attributively in describing the class as a whole, e.g. **A tiger is becoming extinct* (ibid). According to Burton-Roberts (1976), the ungrammaticality of this sentence is due to the fact that the indefinite article only conveys a ‘property generic’. This property is merely accidental and not inherent in the class denoted by the nominal. The zero article identifies the whole class. When used with count nouns, plurality is typically indicated. The definite article is the least flexible of all of them. When used with a singular count noun, it can represent a typical member (see 1.3.4). However, as Chesterman (1991/2005) and Burton-Roberts (1981) point out, when used with plural count nouns, *the* can sometimes indicate a generic quality to apply to less than the whole population.

Some languages do not require generics to be definite. As for English, generics that are definite can mean something slightly different from those that are not (see (50) – (52) and 1.3.4).

2.5 Unifying accounts of definiteness

Despite the inconclusive nature of the research on definiteness, attempts at a unifying account of this phenomenon have been made. Four of them will be described here: Abbott (2004), Huebner (1983), Lyons (1999) and Chesterman (1991/2005). They either represent more recent works on definiteness or are commonly used by other researchers in their studies .

2.5.1 Abbott (2004)

Abbott (2004) is a major proponent of the uniqueness theory. In this analysis, she states her view that all definite NPs are firstly unique. For her, Russell's idea (incorporating inclusiveness) can be formalised as follows (2004: 2):

- a. the F(s) is/are G;
- b. $F - G = \emptyset$.

Thus: “(b) requires the set of Fs to be a subset of Gs. Then for singular definites F is presupposed to denote a singleton set, and for plural definites it is presupposed to denote a set of more than one member” (ibid). The clarity and precision of this formulation, Abbott argues, should be at the core of the meaning of definiteness. However it needs to be enriched/refined by the pragmatic context, in the sense of P-sets *à la* Hawkins (1991). Arguing that three previous attempts at this refinement (i.e. Ariel,

1988; 1990; Gundel et al., 1993; Birner and Ward, 1998) are not adequate in their over-reliance on the notion of identifiability (cf. the confusing treatment of uniqueness and identifiability and the operating assumption that identifiability is achieved by the referent being unique in the context of utterance in 2.2.4). As pointed out above, there are cases in which true identifiability of a given referent is simply impossible.

As these previous attempts at refinement rely on the hearer's role in achieving identifiability and in view of the fact that such identifiability is not always successful, Abbott proposes that the identifying ability of the hearer should be seen as the basis for a conversational implicature (2004:6). Further pragmatic meaning of a definite NP must also be derived from the existence of a P-set.

To illustrate, Abbott (2004: 8 – 13) discusses cases where uniqueness fails to establish the identity of a referent:

53. Yesterday *the dog* got into a fight with another dog. (Attributed to McCawley by Lewis, 1979:348)
54. Switch *the light* on. (Attributed to Löbner, 1985)
55. Waiter, I demand to see *the menu*. (Attributed to Epstein, 1999)
56. The boy scribbled on *the living room wall*. (Attributed to Du Bois, 1980; also 35a)
57. The contestant gave *the wrong answer* and had to be disqualified.

For (53), she argues the definite *the dog* belongs to “a small class of items such as pets, babies, houses [and] vehicles” that are commonly owned by people. As such, these

items have a special relationship with the owners, resulting in their being referred to as, e.g., the owner's own dog. This implies uniqueness. She calls such items "proprietary" (p. 8).

In the case of (54), Abbott argues that the definite marking can be attributed to the fact that items (referents) such as *light* were traditionally unique. This is based on the observation that devices or entities such as lights, elevators, maids and so on operated or were operated, for a significant period of time, as single units. Conventionalisation may play a part, too. Hence British *hospital* as opposed to American *the hospital* (although see our own explanation in 8.2.2).

The NP in (55) is a product of writing like "books, poems, the morning paper and so on". According to Abbott, "the types of these linguistic products are of much greater importance than the tokens, and the [identifiability] signalled by the singular definite is undoubtedly licensed by this common fact" (p. 10).

Unlike Du Bois' 'curiosity principle' (see 2.2.4), Abbott's explanation of (56) is based on the assumption that the NP specifies a location. This is also true for the cases of (35a) and (36). Sentence (57) is the most problematic case. Abbott notes that there can be more than one wrong answer in the context but it is puzzling as to why in (57) the NP is marked for definiteness. Based on sentences like *There was the wrong address*

written on the envelope, Abbott hypothesises that *the wrong* NP is indicative of a location like (56). For her, it also suggests that such NPs are a kind of false definite.¹⁸

Overall, Abbott is convinced that the uniqueness theory is capable of accounting for all uses of the definite article. Uniqueness as the basic meaning of definiteness, however, has to be considerably refined by pragmatics.

2.5.2 Huebner (1983)

Huebner's approach is based on Bickerton's (1981) semantic wheel for noun phrase reference, marked by the features [\pm Specific Referent (\pm SR)] and [\pm Assumed Known to the Hearer (\pm HK)]. It can be seen that the two features are associated with specificity and familiarity respectively. Many SLA studies, e.g. Huebner (1983), Master (1987), Parrish (1987), Tarone & Parrish (1988), Thomas (1989), Young (1996), etc. made use of this distinction to explain learners' stages of article acquisition. Table 2.2 summarises the classification of articles based on these features.

It is easy to see that the table is not comprehensive in its coverage. For example it does not consider idioms or the use of zero in definite NP such as '*I go to school everyday*'. There is also no mention of non-specific *the* arising from [$-$ SR, $-$ HK] as in (48b), reproduced here:

¹⁸ *The wrong answer* may also be conceived as a representative of all wrong answers, a concept not unlike *genericity* or *type*.

48 b. I'm going to have lunch with the president tomorrow – that is, if the election takes place today and we have a president.

However these are not the only criticisms of Huebner's framework. Arguing that [\pm Specific Referent] is not identical to wide scope, i.e. asserting the existence of a referent, Ionin et al. claim that there is a distinction between what Huebner identifies as specific reference and referentiality itself (2003: 151-152; 2004: 9). [\pm Specific Referent] as identified by Huebner is in fact equivalent to a *de re* reading. Recall from 2.3 that specificity is speaker intent to refer. Ionin et al. also append to this definition of specificity the notion of noteworthiness (2004: 7). They argue that if there is something noteworthy about the referent and the speaker intends to refer to it, then the NP amounts to *de re*/ specific reference (or Huebner's [+Specific Referent]), e.g. 'Adam talked to *a girl* - but he wasn't interested in her at all' (cf. *this* in 2.3). When there is nothing noteworthy about the referent and the speaker does not wish to refer to it, the NP is given a *de dicto* reading (Huebner's [-Specific Referent]), e.g. 'Adam would like to talk to *a girl*- it doesn't matter which one '.

Conversely, referentiality has to do with whether or not an entity refers to anything at all (see 2.3). Referentiality does not contrast with non-specificity (*de dicto*) but with non-referentiality. Fodor and Sag call it "quantificational" (1982: 63). This contrast can be seen in their examples:

58. A man just proposed to me in the orangery (though I'm much too embarrassed to tell you who it was).

59. A man is in the women's bathroom (but I haven't dared to go in to see who it is).

The *man* is interpreted slightly differently in each sentence although it is *de re* in (58) and (59). The *man* is referential in (58) but quantificational in (59). This distinction is not incorporated in Huebner's framework.

Table 2.2: Huebner's environments for the appearance of *a/an*, *the* and zero (adapted from Thomas, 1989)

Features	Environment	Articles	Example
[-SR, +HK]	Generic nouns	<i>a/an, the, Ø</i>	Coconut trees are very useful The monkey is a curious animal A camel can go without water for months
[-SR, -HK]	Nonreferential nouns attributive definites non-specific indefinites (etc.)	<i>a, Ø</i>	Chan is a scientist He's looking for a wife
[+SR, -HK]	Referential indefinites first-mention nouns	<i>a, Ø</i>	I live next to a scientist (The scientist keeps ...)
[+SR, +HK]	Referential definites previous mention specified by entailment specified by definition unique in all contexts unique in a given context (etc.)	<i>the</i>	(I live next to a scientist) The scientist keeps to himself though I bought a book but the pages were missing the last person to leave, the same thing the sun In Australia: the Queen, the Prime Minister

2.5.3 Lyons (1999)

Lyons investigates definiteness from a cross-linguistic perspective. Surveying languages both with and without articles, he concludes:

[definiteness] is, as an element of discourse organization, to do with whether or not a referent is familiar or already established in the discourse – thus identifiability rather than inclusiveness. This assumption is bolstered by the observation that demonstratives, which cannot be characterized as inclusive, are invariably treated as definite in interpretation in (in)definiteness effect contexts. (1999:278)

In other words, definiteness can be seen as a grammaticalisation of identifiability. However, given the broad discrepancies of meaning observed in various NPs (many of which are listed above), Lyons proposes that definiteness be seen as a grammatical category, not as a semantic/pragmatic category. This conceptualisation allows for such variability considering that other grammatical categories such as tense and aspect hardly exhibit neat one-to-one correspondence between form and meaning.

Lyons (1999:298-301) modifies the DP structure by postulating that it is in fact definiteness that is the functional category. Given this modification, the DP is actually a definiteness phrase (as opposed to a determiner phrase). Definite articles project a DP and free form articles like *the* occupies the specifier slot. Indefinite articles project a lower-level cardinality phrase and are not regarded as associated with (in)definiteness. It expresses indefiniteness by virtue of the non-occurrence of the definite article. The indefinite article *a/an* occupies the specifier position of the cardinality phrase (cf. Klinge, 2006). According to Trenkic, “[w]hile definite expressions are not incompatible with the concept of cardinality, some phonological constraints do not allow [the] two

articles ...to introduce a single nominal phrase” (2004:1422). Accordingly, a cardinality phrase is maximally projected at a higher node which also dominates a DP. This is the KP (K for case).

While Lyon’s theory is able to account for a number of article-related syntactic phenomena in various languages, it runs counter to the position that I have outlined in 1.2 in which articles are a *class* of (in)definiteness markers. However, an important aspect of this study is language *use* and in that regard, Lyons writes:

A distinction is sometimes drawn between semantic and grammatical definiteness, and I have suggested that generics, for example, may be formally or grammatically indefinite in some languages and yet behave in certain ways like definites because they are in some sense semantically (or, better, pragmatically) definite. More generally, the fact of the referent of a noun phrase being familiar, identifiable, or inclusive may not have any formal place in the grammar of a given language; but it may play an important role in language use, perhaps as part of a larger concept of prominence¹⁹, and function as a trigger for certain grammatical choices. But then there is the grammatical category of definiteness – the grammaticalization of the pragmatic concept, but distinct from it [.]
(1999:226)

Lyons’ theory is a very important one for it explains the lack of uniformity in regard to the forms of definiteness. However it remains that the meaning of definiteness as it is used in a language (i.e. semantic/pragmatic definiteness) is just as important and requires an examination of its own and hence, this study of articles.

¹⁹ Prominence here roughly refers to importance in discourse (Lyons 1999:226).

2.5.4 Chesterman (1991/2005)

Chesterman's theory is based on Christophersen's notion of familiarity. The starting point of his investigation was not the opposition between *the* and *a*, as commonly was the case. He was interested instead in the difference between 'article' and 'no-article' on the surface, which is a line of research in the tradition of the French linguist Guillaume. Following Hawkins (1978), he proposes that the use of *the* signals locatability and inclusiveness. But the presence of articles at all (definite or otherwise) also means that there is a signal of 'extensivity'. To oversimplify, extensivity refers to an entity's maximum potential quality (as opposed to quantity) and hence, generality. A further distinction needs to be made between zero and null articles (see 1.2 for a grammaticalisation account of the two). The zero article occurs before mass and plural nouns while the null article occurs before proper names and NPs such as 'Come along, \emptyset boy', 'He's in \emptyset prison', ' \emptyset Breakfast is ready', etc. (1991/2005:45-47). A distinguishing feature between the two is that the null article cannot occur with a restrictive relative clause (cf. * 'He's not *John* I used to know' and * '*Breakfast* you asked for is ready') unless a definite article is supplied prenominally.

In 1.2 it was argued that the state of no-article before proper names should not be classified as an article at all. This is because the absence of an article here does not mark indefiniteness and the proper name itself is already definite and thus no further marking is required. How do we reconcile this position with Chesterman's argument concerning the null article? According to Payne, "[p]roper names are used to refer to specific individuals both speaker and hearer can identify, therefore they do not usually

appear with articles, modifiers, possessors, relative clauses, or other devices that render nouns more identifiable” (1997/2006: 39). But we have seen that when modified, proper names do take an article (see ‘*the John she used to love*’ and ‘*a stronger Asia*’ in 1.2). Once modified, the semantic constraint of definiteness on the proper name is removed and this is the reason why one of the surface articles has to move in to occupy the available determiner slot in the emerging NP (see 1.2). But there is a way in which we can acknowledge the absence of an article (and its contingent definiteness marking) in a proper name and exploit the null article’s qualities (see table 2.3): this is to rely on the meaning of its name i.e. *null*. Taken to mean ‘valueless’, the null article is equivalent to an article that has no value – definite or otherwise. Its role is solely to contrast the absence of surface marking for proper names with the zero article which, on the contrary, is the state of no-article with the value of being indefinite and is used with other kinds of NPs.

Our discussion so far has not addressed the other NPs that take the null article, e.g. ‘ \emptyset *Dinner* is served’ and ‘You are no match for me, \emptyset *boy!*’ Are they also proper names? Do they behave in similar ways? If we take into account the three features of locatibility, inclusiveness and extensivity, the statuses of these NPs and proper names are the same. Recall from 2.1.3 in which Biber et al. note that the absence of articles in sentences such as ‘We go to \emptyset *church* sometimes’ may represent “the neutralization of article distinctions” (1999: 261) because the referent and the practices associated with it have been institutionalised (Quirk et al., 1985: 277). We may now reinterpret their

observation in Chesterman's null article terms. Notice also their examples from (a) through to (i) in 2.1.3 are not of the zero article but the null one.

The, a(n), zero and null are distributed across the three oppositions of locatability, inclusiveness and extensivity in the following way:

Table 2.3: Distribution of articles across Chesterman's oppositions constituting definiteness (1991/2005: 68)

	Locatable	Inclusive	Limited extensivity
zero	-	±	-
a	±	-	+
the	+	+	+
null	+	+	-

2.5.5 Preliminary conclusion

Uniqueness, as argued by Abbott, proves to be essential for interpreting many definite NPs. However Lyons' observation that the definite article tends to behave like a marker of identifiability is also a useful one. More importantly, following Lyons, we recognise that there is grammatical definiteness on the one hand, and semantic/pragmatic definiteness on the other. Our focus is on the latter, and given that the data will contain not only substitution but also omission of articles, Chesterman's theory regarding zero

and null articles, as well as the notion of extensivity, will prove to be useful for our analysis in chapter 5 onward.

2.6 Some previous SLA studies on articles

The purpose of this section is to discuss briefly the findings of some relevant studies on the acquisition of articles by L2 learners in the light of the theories discussed thus far. Another objective is to enable a comparison of the findings with those of this study.

A major agenda in SLA research is to identify the influence of L1, if any, on the acquisition of articles in L2. Researchers often invoke Universal Grammar (UG) to distinguish the linguistic knowledge that learners have access to either through L1 only or both L1 and L2. On the one hand, there are those who postulate that L2 learners do not have access to UG and their syntactic knowledge of articles is, *ipso facto*, deficient (e.g. R. Hawkins, 2001 and R. Hawkins and Chan, 1997). This belief appears to explain why learners with different L1s are not equally successful in mastering a grammatical form (Trenkic, 2007: 291). On the other hand, others argue that learners have access to UG but this (syntactic) knowledge is not always successfully manifested in terms of morphology on the surface (e.g. Lardiere, 2003 and Ionin et al., 2004). This line of research seems successful in explaining why the variability in learners' production is rarely random.

For example, using the framework in table 2.2, Huebner (1983) and Master (1987) found that their subjects relied on the feature [+Hearer Knowledge] and not [+Specific

Referent] to use *the* correctly. This was based on their findings in which subjects overwhelmingly used *the* in [+SR+HK] and [-SR+HK] contexts (see table 2.2). Thomas, on the other hand, found that L2 learners “associate *the* with the feature [+SR]” (1989: 351). Her evidence is learners’ overgeneralisation of *the* into [+HR] contexts, at least in the initial stage of their acquisition.

More recently Ionin and Wexler (2003) argue that it is not [+SR] that learners associate or dissociate *the* with; it is [+Referentiality] (see 2.5.2). Later, they recast their argument in terms of [+Specificity], highlighting the role of ‘speaker intent to refer’ as the defining element of this reformulated feature (Ionin et al., 2004). To explain the common behaviour of *the* in most SLA trajectories, they propose the Fluctuation Hypothesis where learners “are predicted to fluctuate between the two settings of [Article Choice Parameter] until they are exposed to sufficient input to set the parameter correctly” (Zdorenko and Paradis, 2008: 230). The two settings involved are [+Specificity] and [+Definiteness].

It would be interesting to find out if and how the feature of [+Specificity] influences article use in Outer Circle varieties insofar as it is permitted by corpus data. It must be noted that SLA studies often gather data in non-naturalistic settings and this may have an impact on the production of articles (although Huebner’s longitudinal study is an exception). More discussion on elicited vs. corpus data is found in chapter four.

2.7 Summary

Definiteness viewed in a narrow sense is that marked by articles. Syntactically, articles are determinatives carrying out the function of determiner. The class of articles is constituted by *the*, *a(n)* and zero. Null is not an article but is a useful concept to contrast with the absence of articles as marked by zero.²⁰ Research on the meaning of definiteness can be formal, cognitive-based or both. Russell's notion of uniqueness captures the meaning of a lot of definite NPs marked by *the* but leave very many unexplained. It struggles with the problems of presupposition, non-referentiality and incomplete descriptions. Some solutions have been offered but these have not resolved all the complications. Hawkins' concept of inclusiveness significantly enhances the applicability of uniqueness. A competing theory is Christophersen's familiarity. Making explicit the roles of the speaker and hearer, it accounts for many definite NPs left unexplained by uniqueness. However it too has its limitations. Linguists sympathetic to the concept of familiarity have developed it further into identifiability. The latter is more sophisticated and much broader in its conceptualisation, and thus more effective. However, it is sometimes confused with uniqueness as evidenced by the term 'unique identifiability'.

Studies have repeatedly shown that neither uniqueness nor identifiability is adequate to capture all uses of *the*. It is possible that this is due to definiteness being a grammatical category more than it is a semantic/pragmatic one. Nevertheless more recent accounts have begun to reconcile the two though some linguists are firm in their intention to

²⁰ However, following Chesterman (1991/2005) and for convenience, we will continue to refer to null as the null article in this thesis.

keep the semantically inclined uniqueness and pragmatically inclined identifiability separate. Taking into account the data to be analysed in this study, we plan to adopt a combination approach as well.

Some SLA researchers who believe that L2 learners have access to UG have found that [+Hearer Knowledge] is associated with the production of *the*. Others found contrary evidence and argued that *the* is influenced by [+Specificity]. The influence of this semantic feature on the use of articles in Outer Circle varieties is yet to be seen.

The literature reviewed in this chapter should place us in a good position to interpret the findings of the study and to design the appropriate research methodology. This will be discussed fully in chapter four. The next chapter discusses another dimension of the phenomenon that this study is investigating. This is nativisation.

Chapter 3 Nativisation

This chapter provides an overview of nativisation, the other important component of this study. To understand this linguistic process, the study of language change and variation is described in the next section. Internal and external factors affecting nativisation will be surveyed and the putative division of labour between them is examined in the following two sections. Aspects of nativisation involving phonology, lexis, and grammar (especially articles) are briefly discussed and exemplified in 3.1.3. Section 3.2 explores the topic of *world Englishes* where socio-political aspects of the study of language variation are highlighted and where nativisation plays a crucial role in the characterisation of the Outer Circle varieties. Finally a broad socio-historic description of the four Outer Circle varieties and three Inner Circle varieties in this study is presented.

3.1 Language change and variation

The topic of language change may imply a diachronic perspective that is inconsistent with the aims of this study. However the discussion of language change here has two related objectives. One is to demonstrate the interrelationship between language variation and nativisation and how they both are intricately related to language change (Schneider 2007: 97), for which recourse to diachrony is only natural. Second, it is to acknowledge the fact that any synchronic study will benefit from diachronic considerations. As McMahon puts it, “although historical investigation may be subsequent to synchronic analysis, since it involves the comparison of successive

synchronic states, an adequate linguistic theory must involve a diachronic dimension, and synchrony and diachrony are intertwined” (1994: 11). Undoubtedly the reminder is useful for this study but it will be most apparent when considering the evolution of the varieties under study from earlier stages of British English. There will be a diachronic perspective integrated into the sections containing that discussion.

To say that language changes is perhaps stating the obvious. We can witness the change in the vast differences between Old English, Middle English and Modern English, or in the less conspicuous dissimilarities between British English and Australian English, or even in the speeches of the old and young in the same community. These changes are found in such separate domains of language as grammar, phonology, semantics, etc. In fact, changes in language are pervasive; compare the speeches or written texts of any two social groups and chances are there will be some systematic differences (the term ‘idiolect’ even captures this idea at the individual level).²¹ But what drives these changes? Linguists find it hard to agree on the answer. Early researchers from the Neogrammarians, through the Structuralists, to the Generative grammarians have focused exclusively on the internal properties of language (McMahon, 1994; Farrar and Jones, 2002). For example, emphasising regularity in linguistic phenomena (in common with the view espoused by the dominant Darwinism at that time that much of the world itself is regular and explainable), the Neogrammarians regarded sound changes as “mechanical and physiologically motivated” (McMahon, 1994: 21). This view precluded any consideration that could be made of external factors such as language

²¹ In fact the choice of the word *social* here implies that a lot more will be said about society and its significance for language.

contact or power relations and instead focused exclusively on parameters such ease of articulation and so on.

Working on the premise that the dichotomy of internal and external factors motivating language change is essentially flawed because language change is just too complex, Farrar and Jones (2002: 2-3) argue that historical linguistics has not yet fully embraced factors that are otherwise identified as ‘external’ to explain the phenomena it studies. This indifference is being nurtured by what they call the “either-or” mentality and the “if-in-doubt-do-without” mentality. The first “demands that [linguists] make a decision between internal and external motivations when seeking to explain language change [but]...[i]n practice, when such a decision is made, internal factors are considered in some sense superior and external factors usually lose out” (2002: 4). The “if-in-doubt-do-without” mentality dictates that “[e]xamining whether [external factors play] a role in change is ... seen as a last resort, and *if in doubt* [linguists] should *do without* and simply not take this final step” (ibid.). Of course there is a danger in continuing to adopt either mentality. Foremost of all, we could not arrive at an understanding of how Outer Circle varieties came into being, and explain the process of nativisation itself. This “dangerous” dichotomy (Farrar and Jones, 2002: 1-3) of external and internal factors will be examined again below in the discussion of diffusion.

In a study of what he calls “postcolonial varieties of English”, Schneider (2007: 97) considers language change as the underlying process that has shaped the varieties into what they are. Obviously not subscribing to the two mentalities described above,

Schneider attributes many of the varieties' unique forms to English (as spoken by its native speakers from Britain at the time) coming into contact with the indigenous languages spoken by inhabitants of the foreign soil where the British speakers became settlers. This is not to say that internal factors played no role. In fact, Schneider proposes an elaborate model (see figure 3.1 below) to explain the intricate interrelationships between internal and external factors, hypothesising that “innovations occur in the speech of individuals because of internal conditions but they spread to the community for external reasons” (2007: 99). We will return to this hypothesis when the process of diffusion is discussed below. Prior to that discussion, there is a need to examine language change using a metaphor that is increasingly used to describe it – evolution.

In an effort to build a parallel between biological evolution and language evolution (see McMahon, 1994: 314-315 on the benefits of this exercise), Rosenbach (2008) proposes a view of language change as cultural revolution “with social, communicative, and psychological factors determining language change rather than evolutionary forces known from biological evolution” (p. 24). One of the properties of biological evolution that are being borrowed to explain phenomena in language change is ‘altered replication’, which is said to cause variation (much in the same way DNA is gradually altered to give rise to mutation). Rosenbach (2008: 33) points out, however, that how exactly an alteration occurs and how it is replicated in language change are not fully known. Correspondingly, she further argues that innovation, i.e. where alteration takes place, and diffusion, i.e. where replication occurs, are two separate processes, unlike

how they are frequently presented in the literature, i.e. conflated. Innovation will be discussed in some detail first before diffusion is presented.

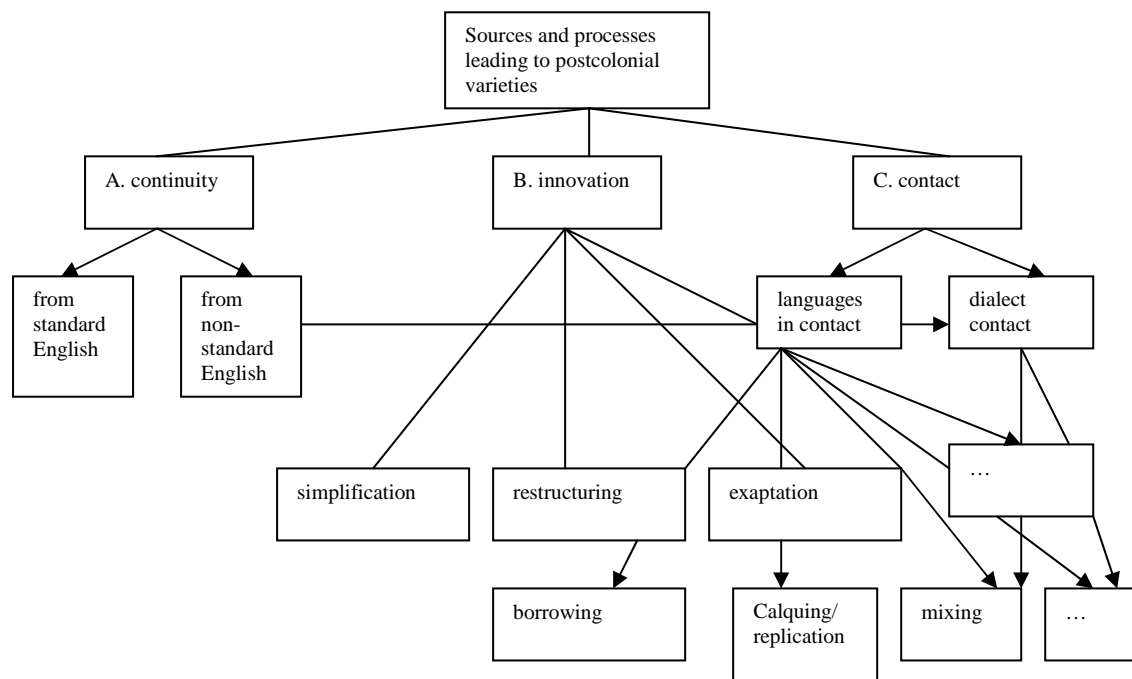
3. 1. 1 Innovation

Where innovation is concerned, Rosenbach lists three triggering mechanisms – exaptation, analogy and reanalysis. ‘Exaptation’ is likened to recycling of linguistic materials already in use for new purposes. ‘Analogy’ or analogical extension is “the generalisation of a form/construction to a new context or to new forms/constructions” (p. 35). ‘Reanalysis’ is the creation of a new form/construction altogether (p. 36) where “subsequent speaker generations analyze and understand the same constituent sequences differently in their mental grammars, [leading] to further modifications and realignments” (Schneider, 2007: 105). Given that all these processes are characterised as occurring speaker-internally, it is no wonder that Schneider considers innovation as internal. For Schneider, innovation is a central part of ‘nativisation’ (2007: 71 -112). Structural nativisation, which he refers as “the emergence of locally distinctive linguistic forms and structures” (p. 71) is complex in origin and its explanation requires that the interaction between external and internal factors be made explicit. A more extended discussion on nativisation is given in 3.1.3.

Figure 3.1 thus presents his own model describing “sources and processes leading to postcolonial varieties of English” where innovation is only one of three major sources of variation. Schneider reminds us that the empty cells in the model “[suggest] that the listings are not exhaustive, and the arrangement is definitely not the only possible one;

the [model] just suggests one way of systematizing a set of complex interrelationships” (2007: 100). However if we continue to adopt the notion of altered replication (in evolution) to explain language change (ibid.), it will be seen that ‘continuity’, i.e. “direct historical transmission across generations” from the source/mother variety, lies outside our concern at the moment. It will, however, be discussed in the broad socio-historic descriptions of the varieties below. In addition, in following Schneider’s own advice to modify the model to suit one’s needs, it would require less complication to interpret the model if ‘innovation’ were not placed on par with ‘contact’. For our purposes, it would be helpful to assume that all language contact processes were forms of innovation, placing ‘contact’ itself above ‘innovation’.

Figure 3.1: Sources and processes leading to postcolonial varieties (from Schneider 2007:100)



Obviously some processes have been given different names or subsumed under a different category by different writers and this is admitted by both Rosenbach (2008) and Schneider (2007). Schneider himself does not explain all the processes he includes in the model. The following are the processes that are identified in either Schneider or Rosenbach (as both are most recent works on language change and nativisation and include in their surveys most other significant previous works).

In addition to exaptation, analogy (or analogical extension) and reanalysis identified by Rosenbach (2008) above, Schneider lists the following (pp. 102 -107, all examples from this publication though not necessarily his):

1. simplification

e.g. East African short vowels in contrast to RP's more elaborate six short vowels, six long vowels and two unstressed vowels.

Schneider contends that simplification is a misnomer in that having to do with a reduced set means a lot more creativity is required of speakers to serve communicative needs (p. 102). For him, 'analogy' is a mechanism/sub-process of simplification. Others include 'regularisation' (e.g. application of *-ed* in 'knowed') and loss (e.g. omission of inflectional past tense marking in Australian Aboriginal English).

2. reduction of markedness

e.g. the use of 'knowed' instead of 'knew' is also an instance of this process.

3. restructuring

e.g. non-observance of inversion in *wh*-questions such as 'What this is made of?' and observance of it in indirect questions such as 'I asked him where is he?' in Pakistani English. This "involves the systematic rearrangement and reinterpretation of constituents and constituent sequences" (p. 105) which according to Mufwene (2001: 13) includes "loss of some units or rules, addition of new ones, and certainly modifications in the direction of simplification, generalization, or complexification by the addition of conditions to the application of a rule." Schneider points out that this term may well be "a superordinate term of grammatical change which subsumes simplification, exaptation, and so on" (p. 105). For Schneider as well, 'reanalysis' is a sub-process of restructuring.

4. grammaticalisation

e.g. the use of *know* as a discourse particle in utterances like ‘The coffee is hot, know’ in Singaporean English. The origin is the more common English discourse marker ‘you know’.

Calquing or replication is discussed by Schneider (pp. 107 -108) specifically under the parameter of contact, although according to his model, he allows it to be subsumed under exaptation. Overall we find that many of these processes overlap and, in some instances, are really two sides of the same coin. Although this potentially confusing treatment of innovation was mentioned earlier, we have to agree with Farrar and Jones (2002: 1-2, see above) that the external/internal dichotomy is a “dangerous” one. While it may help to identify certain phenomena by specific names, a researcher cannot let himself/herself be restricted to the limitations that may be inherent in them. Having said that, our position is that both internal and external processes/factors can be labelled as they are only when the context of each variable they seek to explain is considered. In this study where nativisation is a central construct, it will be seen that some of them can be safely called ‘internal’ or ‘external’, while some others cannot.

3. 1. 2 Diffusion/Spread

Once there are two variants or more (following an instance of innovation or more), how does a community of speakers come to use the one that they do collectively? This topic is also discussed under ‘selection’ (Rosenbach, 2008: Croft, 2000). As noted by

Rosenbach (p. 40), variation can lead to the loss of at least one variant or specialisation of each variant according to contexts and speaker groups. She elaborates:

Selection operates whenever we find differential replication of variants, i.e. whenever there is a shift in the distribution of the two variants. That is, once variants have settled down in their respective functional and/or social niches and have thus reached a stable distribution or equilibrium, selection no longer comes into play. However, selection is present in the process leading to functional/social differentiation or whenever variants come to be associated with new values. Under this view, selection does not necessarily reduce variation (as is the commonly held view) but may also be present in any process that changes the distribution of variants.

(2008: 40)

Naturally there have to be factors to determine the selection but unfortunately their research is inconclusive. Earlier on, Schneider's hypothesis that innovation starts internally in an individual and spreads externally to others was presented. This will be examined now.

Schneider's work (2007) is greatly influenced by Mufwene's (2001) *The Ecology of Language Evolution* in which the latter seeks to explain the evolution of pidgins and creoles (and other kinds of varieties). Central to Mufwene's notion of ecology is a 'feature pool' that receives the features of all the languages that come into contact with the native language(s) of a speech community. Ecological processes, much like those in biogenetics, are those that govern the linguistic outcomes of the interaction between the rich and varied features. This means that there is a 'founder effect' that establishes the form of a variety. In the case of the varieties under study, this refers to English. Note that the feature pool also implies the contact of different dialects of English, indigenous languages and other languages introduced to the setting. However, as in biogenetics,

much is unknown about the precise selection of features in a variety formed from those inputs. Mufwene explicitly leaves this part of the equation to chance: “ecology rolls the dice” (2001: 145ff).

It is easy to appreciate Schneider’s hypothesis once the role of language contact in Mufwene’s notion of ecology in language change is understood. It appears that language contact is enacted by agents external to languages themselves (i.e. speakers) but innovation (i.e. alteration) originates *in* a speaker and spreads from one speaker to another. Although Schneider does acknowledge the significance of so-called internal factors by stating that selection includes, among others, “the nature of the linguistic input elements, surface similarities and typological degrees of relatedness between the languages involved” (p. 23), the neat characterisation contained in his hypothesis betrays the actual complexity of this process. This recalls the danger of the internal/external dichotomy discussed in the previous section. Mufwene clarifies:

When ecology is adequately factored in, these questions boil down to the problem of ‘multiple articulation of selection’ in a population, with different selections applying concurrently at different levels within the population. On the one hand, these are selections made by individual speakers which assign each idiolect what in biological terms would be identified as its genotype. Each idiolect has its own idiosyncracies, despite overwhelming similarities with others in the same communal language. On the other hand, the community at large makes its own selections through the fact that the innovations or idiosyncracies of some speakers...are copied by other speakers while others are not[.] The community-level selection is what produces macroevolutionary developments identified as changes in a communal language. *However, so far the principles regulating both individual and group selections are not fully understood.*

(2001: 17-18, emphasis added)

Similarly, Rosenbach disputes Croft's claim (2000) that "it is social factors – and *only* social factors – that drive the selection process" (2008: 42, emphasis in original). Croft argues that innovation, a process prior to selection, is functionally motivated, i.e. alteration arises out of socially communicative needs such as accommodation (resembling one's speech to others') and observing social statuses among speakers. It follows that selection is functionally based and thus, socially based too. Citing findings from other studies Rosenbach shows that after the appearance of several variants of the same linguistic item, which speakers initially treat as free variation, selection is made based on functional factors and these are not necessarily socially based (pp. 45 – 46). Rosenbach summarises thus: "[u]ndoubtedly, social factors play an important role in the promotion or demotion of variants, too, as diachronic sociolinguistic research has shown...but the evidence available does not speak for the exclusive role of social factors in the selection process" (p. 44). Furthermore Rosenbach echoes the concern of Farrar and Jones (2002) in that what distinguishes functional (i.e. internal factors) and social (i.e. external factors) selection is "a matter of personal preference (or theoretical orientation)" (p. 47).

Schneider's (incomplete) list of factors contributing to the selection process is given below, where he writes, "it may be hypothesized that ... the external ones are more powerful than internal constraints" (2007: 110 -112). It would be best to interpret Schneider's statement and the following factors in the light of the above discussion.

1. demography

A higher number of speakers using a form will most likely ensure its continuing use.

2. frequency

Related to demography, frequency refers to the higher number of occurrences of a form bringing about a deeper cognitive entrenchment and ensuring its continuing use.

3. historical depth

The earlier a form is available (chronologically) for the speech community, the more likely it will survive.

4. markedness

Unmarked forms, with more natural and more common features, have a greater chance of being selected.

5. salience of forms

A higher degree of salience (i.e. “more perceptually and cognitively prominent” (Kerswill and Williams, 2002: 81) are more likely to be selected. Salience, however, may conflict with markedness.

6. transparency and regularity

These are related to markedness. Regularity and transparency in patterns ensure more likely selection.

7. status of speakers

Status refers to prestige accorded to a speaker. Often forms used by a higher-status speaker are selected by others (perhaps lower in status). But social status doesn't always correspond with official status. Forms employed by lower official status speakers can be copied by higher-status speakers for social reasons.

8. identity marking function of linguistic forms

Related to (7), speakers may use forms to forge a certain identity.

9. similarity or difference between L1 and L2

Similar and identical forms in the two languages can facilitate selection. But forms that are so close as to cause confusion may give rise to reanalysis and/or be avoided.

3. 1. 3 Nativisation as a process of language change and variation

Nativisation provides an explicit contact dimension to language change and variation. Although the role of contact in the discussion of the previous two sections was highlighted, notice that if it were factored out, language change and variation processes would still take place, i.e. within a single language. Similarly, contact alone would not lead to nativisation. For instance, the use of English loan words in French (e.g. 'le weekend') does not turn Parisian French into a nativised variety of English. Literature suggests that ultimately there are socio-political motivations underlying what linguists call nativisation. This will be examined towards the end of this section and in the next

section on ‘world Englishes’. Presently we discuss it in terms of language change and variation.

Kachru considers nativisation to be a “linguistic readjustment” where it involves “the approximation of a language to the linguistic and discoursal characteristics of the native (or dominant) language of the area into which it has been transplanted” (1992: 235). This recalls Mufwene’s use of ‘founder’s effect’ mentioned above. According to Lowenberg, the effects of the readjustment must be “stable” and are found “in the forms and functions of a fully elaborated variety of English[.]” (1990: 158). Thus, it is to be contrasted with ‘interlanguage’, which “refers to equally systematic, but more idiosyncratic and less stable, deviations from the norms of any variety of English – native or nonnative – by learners of that variety” (ibid).

In 3.1.1 we reviewed some of the mechanisms that had been proposed for innovation. These include simplification, reduction of markedness, reanalysis and so on. These same mechanisms are likely to be at work if nativisation is brought into the picture. However strict labelling of external vs. internal factors, tempting as it may be, should be avoided at this stage. Some linguists (e.g. Trudgill, 2004) believe that it takes approximately fifty years (three generations since the first contact) to give birth to a colonial variety, or a native English spoken outside Britain (Fennell, 2007). Bilingualism and/or multilingualism notwithstanding, this is probably true for varieties in the Outer Circle too (see 3.2 below).

In 3.1.1 some examples of innovation in forms were given e.g. ‘knowed’ or ‘what this is made of?’ A question that one would immediately ask is where exactly the variation resides. Is it in the addition of the *-ed* suffix, or in the lack of inversion? Or is it in the culturally tied ways of expressing things in English, for example in the opening of Japanese business letters that makes reference to the splendour of cherry blossoms when they are in bloom (e.g. Irving, 1984: 142)? Nativisation, as Lowenberg points out above, occurs in the forms and functions of a variety of English. This particular Japanese way of opening a business letter is thus an instance of nativisation and there is undoubtedly a need to study the ways in which English is used in Outer Circle varieties to carry out such functions in their speech communities.²² While pragmatic studies of this kind offer rich findings about language and culture and their interrelationship, e.g. Nkemleke’s study (2006) of ways to express deference in dissertation acknowledgments and private letters in Cameroonian English, this line of investigation lies beyond the scope of the present study. Our focus here is on form.

In terms of form one can find it most conspicuously in pronunciation and vocabulary (see 3.1.1). In Malaysian English, for example, the word ‘academic’ is stressed on the second syllable and the vowel in it is pushed from the centre to the middle front. Also in the same variety, there is a local word to denote being ignorant or slow to catch up – ‘blur’. There are many other examples of nativised pronunciation and vocabulary in this variety and others. However in their description of Australian English, Collins

²² The term *speech community* is used to refer a loose meaning of a group of speakers of the same variety.

and Peters remark that a distinctive profile of that variety is found in “the interface between grammar and lexis” (2004: 593). This claim is corroborated by Schneider who asserts that such kind of variation is prevalent in nativisation, which is considered one of the phases in his model of postcolonial English development (see 3.1.4). This makes cases like ‘knowed’ and ‘what this is made of?’ above common nativised forms. It will be shown that nativised patterns of article usage also fall in this category.

Accounts of syntactic variation in varieties of English worldwide (Kortmann, 2006 and Kortmann and Szmrecszanyi, 2004) attest to the pervasiveness of article variation.

Kortmann observes that:

This involves either the omission of a definite article (e.g. ‘Father rented the farm under Squire’, ‘Take them to market’) or an indefinite article (e.g. ‘I had nice garden’, ‘They had awful job’), or their insertion (e.g. ‘about a three fields’, ‘about a seven inches square on a board’, ‘I left the school in early age’, ‘Do they keep the goats?’). Another option found in the same or other varieties is the use of the definite article where Standard English uses the indefinite article (e.g. ‘I had the toothache’, ‘he’s the wise boy’). (2006: 606)

A cursory glance at Kortmann’s examples will show that the omission, insertion and exchange of articles are probably not random. Syntactically each NP is still well-formed (see chapter one) providing one is willing to abandon traditional distinctions of countness and massness. Similarly the semantics of definiteness as identifiability is not necessarily compromised if it is (non)identifiability that one looks for in, say, ‘Take them to market’ or ‘I had nice garden’. Of course these claims need substantiation and this will be the focus of chapter six onwards.

Meanwhile, Kortmann and Szmrecszanyi show that “irregular use of articles” (2004: 1155ff) is among the top 15 most frequent morphosyntactic features in the 46 varieties of English they surveyed. It is found in 33 of the varieties, in the regions of British Isles, Caribbean, Australia, Africa and Asia. The varieties representing these regions will be described in 3.3.

To end this section, one important question needs to be asked: what qualifies a variety as nativised? Why isn't Parisian French, upon use of several English loanwords, a nativised variety of English? Perhaps one can quickly point out that the language is French and not English (cf. Mufwene's founder effect). But how do we account for the differences between the English used in Kingston, Jamaica and that in Bangkok, Thailand? Both are spoken by significant numbers of speakers. In both contexts there are good reasons to use English consistently (e.g. tourism, international trade and communication). But whereas Jamaican English is readily categorised as nativised, Thai English is not. Relating this issue directly to innovation, Bangbose writes:

A non-native English situation is basically an innovative situation involving certain well-known processes of nativisation....The main question that arises with innovations is the need to decide when an observed feature of language is indeed an innovation and when it is simply an error. An innovation is seen as acceptable variant, while a error is simply a mistake or uneducated usage.

(1998: 1-2)

Using criteria proposed by Mollin which she uses to assess the eligibility of Euro-English as a “legitimate” second-language variety (2007: 167), it can be shown that all

otherwise Outer Circle varieties selected for this study are indeed nativised. Mollin’s criteria catalogue is given as Table 3.1 below.

Table 3.1 Criteria catalogue for ESL-varieties (Mollin, 2007: 173)

Expansion	<ul style="list-style-type: none"> Extensive bilingualism Use in the domain of education Use in the domain of administration Use in the media Use in creative writing Use as a contact code
Nativisation	<ul style="list-style-type: none"> Extended register and style range Distinctive phonology, lexicon, syntax, discourse style Characteristics must be communal, not idiosyncratic New features must be systematic
Institutionalisation	<ul style="list-style-type: none"> No gap between performance model and linguistic behaviour Acceptance of the local variety and its label Beginning codification and official recognition of the variety

As Mollin explains, “[e]xpansion, nativisation and institutionalization all need to [take] place – in this order (even though there may be overlaps) since institutionalization preresquires nativisation, while nativisation can only take place after or concomitantly with expansion” (2007: 173). Mollin admits that the third of them, institutionalisation, is “the least necessary of all” (ibid.), suggesting that varieties that are still exonormative (i.e. aiming at a native variety’s standard as target) can be nativised.

What is also clear from table 3.1 is that for any variety to fulfil the nativisation subcriteria, a significant level of participation from government institutions in a given speech community (i.e. country) has to be available to initiate and sustain at least some

of the activities of its prerequisite – expansion. In fact, were we to push this argument further, it could be seen that institutionalisation is actually a socio-political strengthening of expansion. Examining, albeit briefly, this socio-political dimension of nativisation is the focus of the next section.

3.2 World Englishes

There are nativised varieties of French in some parts of the world, e.g. Morocco, but very few linguists have been inclined to call them ‘Frenches’, although notice that European French (or the French spoken in France and Switzerland, for example) is usually differentiated from Canadian French. The use of the term ‘Englishes’ is not without controversies. Why use the term at all? One reason is that the vast growth of the number of English speakers around the world has called for a re-evaluation of how English as a native language is conceived. Crystal estimates that there are more than 1 billion of them worldwide (2006). But the significance is not in numbers alone. With numbers, diversity arises. Change, as it is often said, is not always a good thing and objections to the recognition of non-native varieties of English were many (e.g. Quirk, 1985). Kachru argues that defendants of native English standards are ignorant of “socially-realistic linguistics” (1986) where, following the British linguist J. R. Firth, language should be viewed as ‘function’ (p. 26). Sociolinguistic realities such as these require a significant attention be given to language context (p. 28ff). It follows that a pluralistic view of English is needed to provide a framework for the various theoretical and research concerns that arise from this paradigm shift (Bhatt, 2001: 257). It is reassuring too that Bamgbose points out that “[t]oday, few serious scholars...will insist

that a non-native English...is a transitional and unstable code striving for perfection” (1998: 1).

Kachru’s model of three concentric circles (see chapter 1) of World Englishes divides varieties of English into three groups and it was mentioned in chapter 1 that the non-native varieties in this study are essentially those from the Outer Circle countries. I aim to prove that the overlap is merely incidental as there are now varieties that should be formally grouped in the expanding circle but which belong in the Outer Circle precisely because of nativisation. One example is Hong Kong English (cf. Lee and Collins, 2004). In addition, in Schneider’s Dynamic model (2007) where the development of each variety is tracked through five phases of evolution (i.e. foundation, exonormative stabilisation, nativisation, endonormative stabilisation and differentiation), nativisation could sometimes emerge as an endpoint where some Outer Circle varieties stagnate (e.g. Malaysian English and Philippine English) but simply a middle point passed by others on their way to a higher phase (e.g. Singaporean English).

The one thing that all nativised varieties share in common is expansion (see Mollin’s criteria above). While it is also a prerequisite for nativisation, governmental participation crucial for institutionalisation (see the end of the previous section) exists in different degrees and forms for different varieties. In Malaysia, for example, governmental support for institutionalisation has to be tempered with nationalistic aspirations that require the Malay language be given priority in many domains of life. This is no longer true in Singapore where, although Mandarin is considered an

important language, English is guaranteed of its supremacy. By the same token, any variety in the expanding circle could be selected for this study if nativisation has indeed taken place. Japanese English (spoken in Tokyo, perhaps) and Swedish English (spoken in Stockholm) come to mind. Having said this, it must be admitted that the geographical distinctions of Kachru's concentric circles are neat and very convenient to follow.²³

World Englishes, then, serves as the perfect framework for this study. The latter includes as its variables British English, American English, Australian English, from the Inner Circle, and Singaporean English, Indian English, Kenyan English and Philippine English, from the Outer Circle. The framework succinctly sums up the relationship of these varieties in terms of one another. The following section will give a brief socio-historic profile of each of these varieties.

3.3 Describing varieties

An important point that must be made to preface this discussion of British English and all the other varieties is that each of them is not monolithic. Each of these varieties is made of several dialects with different pronunciations, vocabularies and, sometimes,

²³ There is a further qualification that needs to be made regarding nativised varieties. Mufwene (2001) is among those who warn that pidgins and especially creoles are often differentiated from other varieties because of what turn out to be racially tinged considerations. In the face of his arguments concerning the equality of *all* contact languages in the ecology of language evolution, it is the notion of *continuity* (see Schneider's model in figure 3.1) that sufficiently keeps nativised varieties and creoles apart. As pointed out by Fennell, although both types of varieties were subjected to nativisation processes, education to the locals provided by settlers (i.e. continuity) ensured that nativised varieties were significantly closer to the mother variety (2007:195). On the other hand, creoles developed from pidgins and provision of education by settlers did not take place during the early stages of contact. Continuity in this sense then does provide a crucial criterion in characterising nativised varieties.

grammars. However, the variability within one variety must be considered less significant when compared to those of other varieties. It is fortunate that the data from ICE (see chapter five) are collectively regarded as standard varieties (Nelson, 1996) and this fact will definitely facilitate the interpretation of findings later. Descriptions of the varieties here too will generally restrict themselves to the standard ones.

3.3.1 British English

Old English is a West Germanic language brought to Celt-speaking Britain by Germanic peoples from northwest Germany and northern Netherlands. It was more a collection of many dialects than a single identifiable language. Old English was later influenced by the invasion of two other Continental tribes. The first was the Vikings who spoke Old Norse, a language distantly related to Old English itself. While the Vikings made considerable impact on Britain in general, their language - Norse - had more limited influence on English (Startvik and Leech, 2006: 24-25).

The second invasion was by the Normans in the 11th century. They spoke Old Norman, which was a Romance/French language and one that came to influence English in several important ways. Under the rule of the Normans, English began to borrow heavily from French. Middle English is said to begin developing around this time with rapid vocabulary expansion taking place and the complex inflectional system of Old English being reduced and simplified (Startvik and Leech, 2006: 36-37).

The start of Modern English is marked by the spread of English to domains previously occupied by French such as the king's court , the law court and the Church. The propagation of English was also facilitated by a pioneering writer, Geoffrey Chaucer, whose *Canterbury Tales* became a major milestone in English literature, and by an enterprising printer, William Caxton, who printed and disseminated such works.

Later developments, e.g. the popularity of Shakespeare's works, the publication of King James Bible, the vigorous documentation efforts during the Restoration, etc. all helped to transform English into its more modern form (Startvik and Leech, 2006: 46ff). However, it is important to note that the English spoken in Britain nowadays is not the same variety that was transplanted to British colonies. If anything, present-day British English is also a product derived from what was spoken on the British isles a long time ago (Algeo, 2006: 1).

3.3.2 American English

American English (AmE) has a long and varied history. The account given here necessarily glosses over many important details and overlooks a large number of turning points in the history of this variety. AmE began with British colonisation, but it was one that started at least a century earlier than that of other colonies. The first wave of English-speaking settlers arrived in North America in the 17th century. As it was a large continent, settlements were sporadic though concentrated on areas on the east coast (the first to reach across the Atlantic). The arrival of immigrants, motivated by various social conditions in Britain during that period, was intermittent and occurred

over a period of several hundred years (Startvik and Leech, 2006: 78ff). Soon the pursuit of more land meant that contact with the native Americans grew increasingly hostile. Ultimately overcome, indigenous influences on AmE are thus minimal and largely relegated to place names and native flora and fauna.

Early AmE displayed a great deal of variability as the feature pool consisted of several dialects of English and the languages of non-English speaking immigrants and African slaves. As a colony, America clung to the British standard(s) but the transplanted language was already being nativised in phonology, lexis and many aspects of grammar. Gradually loyalty to Britain weakened as various socio-political developments in both the colony and Britain led to its own independence on 4th July 1776. There were rigorous attempts to establish AmE as the language of the new nation, although instances of resistance (in a form of loyalty to British English) were sometimes encountered.

Nativised patterns of AmE were increasingly accepted by the general population in the 19th century as the new nation consolidated itself as an able and independent country. At the start of the 20th century, the American identity was forged and with it, AmE became a major variety of English. Speakers in the vast country gradually emphasised their diversity and dialects of the American variety were increasingly becoming distinct from one another.

3.3.3 Australian English

Regarded as one of the Antipodean varieties, Australian English (AusE) is more recent than AmE in that settlers from Britain only arrived in the country in 1788 (18 years after Captain Cook's exploration and nearly two centuries after colonisation of North America began). The endonormativity of AusE has been argued for by Collins and Peters (2004) who present evidence in its morphology and syntax. The "British base of grammar brought through continuous immigration from Britain and continuous administrative, social and cultural connections" has been weakened by increasing interface with AmE (2004: 607-608). However this does not mean that AusE is drifting closer to its American cousin. First, some grammatical features such as *shall* are so greatly disfavoured that this variety is emerging as charting its own territory. Second, although the Australian evidence sometimes points to an intensification of usages that are also available in BrE and AmE, they remain essentially innovations from the "raw material" of the former two varieties (2004: 607).

Variation in AusE is often described by the cline between Broad, General and Cultivated varieties (Mitchell and Delbridge, 1965). The three levels often correlate with levels of formality and education (Schneider, 2007: 126). Other than this, AusE is often noted for its lack of regional variation. This homogeneity, however, is frequently challenged by speakers of AusE themselves and large numbers of immigrants from continental Europe and Asia in recent times may very well add to the weight of this assertion (Burrige and Kortman, 2004: 548-549).

3.3.4 New Zealand English

James Cook's navigation of New Zealand took place several years earlier than that of Australia. However, English only spread on the islands about a century later when immigrants from the British Isles began to arrive. Unlike Australia, New Zealand was not a penal colony. Thus, its settlers were not convicts but people from "higher social strata or from rural areas in the home country" (Startvik and Leech, 2006: 106).

From the beginning, New Zealand English (NZE) had to recognise the special status of Maori, the native language of the Maori people. The Treaty of Waitangi, signed by tribal leaders and the British Crown in 1840, ensured that a close contact between the two peoples and their languages was maintained. Thus, NZE was impacted by Maori in very significant ways. Place names, vocabulary for flora and fauna and words relating to Maori culture and society are incorporated in its lexicon. Maori was made an official language in 1987, ensuring its importance in New Zealand society.

NZE is also regarded as homogeneous, except for the speech of several Scottish communities on the South Island. As the other major antipodean variety, NZE is frequently noted for its similarity to AusE, much like Canadian English is commonly thought to resemble AmE. However closer inspection will reveal several differences, especially in the sound system.

3.3.5 Indian English

English in India has a long history. It was brought to the country in 1600 following the establishment of the East India Company. Early contact between Indian languages and English took place through trade and missionary activities. This only increased almost two centuries later when India finally became an exploitation colony, mostly through education.²⁴ However English was often resisted by nationalists and thus its development is often viewed as consisting of “paradoxes” (Schneider, 2007: 164). With nationalistic resistance on the one hand and support due to its role in socioeconomic development on the other, English continued to spread in India, albeit in certain circles of the population. It also tended to be reserved for more formal domains. Adding to the adverse condition for its development is the reluctance of the British to allow Indians a full access to the language.

Despite these shortcomings, English has become an integral part of the nation, no less in its role as an official language (cf. the neutrality of English in the case of Singapore). Not all sections of its society are conversant with the language, and obviously this is purely a voluntary outcome for some, but there are signs of it being appropriated by few members from the lower socioeconomic background precisely for socioeconomic purposes. Overall the status of English in India is remarkable given the persistent lack of enthusiasm showed by their colonisers in sharing the language with the indigenous

²⁴ Like in most exploitation colonies (e.g. Malaysia, Singapore, Kenya, Tanzania, etc.) education was not provided as free-for-all. It was carefully restricted, at least initially, with the main aim being to educate a smaller number of indigenous elites who would later serve as mediators between the colonisers and the colonised (Schneider, 2007: 38 and elsewhere, Fennell, 2007:193-194, *inter alia*).

peoples and by many nationalistic Indians for whom the state of affairs persists to this day.

In the domains in which it was being used, English flourished after the country's independence in 1947 (Gupta, 2001 in Schneider, 2007: 165). As a result, nativisation in Indian English (IndE) is quite extensive in pronunciation, vocabulary and grammar. Preference for an endonormative standard has been identified in certain sections of the speech community but given the constant tug-of-war between resistance and acceptance in the general population, it is not certain if this will translate into action in the near future.

3.3.6 Singaporean English

One of the best-documented Outer Circle varieties, Singapore English (SgE) owes its current status to its colonial history (Lim and Foley, 2004:2) but most significantly, the English-based bilingualism policy implemented by its government (Schneider, 2007: 156). Acquired by Stamford Raffles for the British India Company in 1819, the island was not a settlement colony unlike America and Australia. English, as was typical in exploitation colonies (Mufwene, 2001), was introduced mainly through administration and education.

The main ethnic groups of the population – Chinese, Malays and Indians – were later joined by other ethnic groups, e.g. Arabs, other Asians, Europeans, all of which came into contact with English (Schneider, 2007: 154). English developed in tandem with

various economic activities that propelled the island to prosperity. The end of World War II (1945) saw the erosion of colonial power on the island. Following a brief formation of Malaysia with the rest of the Malay states on the Peninsula, Singapore broke away to independence in 1965. Faced with nation-building issues, English was chosen as the unifying language due to its neutrality (while maintaining the official status of other main languages, i.e. Mandarin, Malay and Tamil).

Present-day Singaporean English is widely used in various domains of life. Singaporeans are said to be “English-knowing bilinguals” (Pakir, 1991) with knowledge of English and a mother tongue (Lim and Foley, 2004: 5), although the status of the latter for many younger Singaporeans is increasingly disputed (e.g. Foley, 1998). Widespread use has also resulted in a cline of bilingualism (Pakir, 1991) that also correlates with formality and education. A colloquial variety, Singlish, is widely used in informal contexts and has become a marker of identity for many Singaporeans (Schneider, 2007: 158).

Endonormative tendencies are evident in both the countrymen’s willingness to uphold their own standard and in some aspects of SgE especially in pronunciation and vocabulary (Schneider, 2007: 159). Institutionalisation, however, has not fully taken place especially since evidence of the complaint tradition (Milroy and Milroy, 1999: 24ff) can occasionally be found.

3.3.7 Kenyan English

As part of East Africa, Kenyan English (KenE) is frequently treated as part of a supposedly homogeneous language called East African English. However postcolonial socio-political developments in each of the East African nations – Uganda, Tanzania and Kenya – have pulled each variety in different directions (Schneider, 2007: 189). Uganda (ibid.) and Tanzania (ibid. and Hudson-Eattle and Schmied, 1999:1) have restricted the development of English after independence while Kenya “has opened itself to the world and has seen the English language thriving, within limitations” (Schneider, ibid.).

Although contact with Kenya was established a century earlier, British interest in the country only developed significantly in 1902 (or after the construction of a railway providing access and mobility for expansion of economic activities between the coast and the hinterlands) (Schneider, 2007: 189). Many British (and other European) settlers arrived later to exploit the agricultural and other economic opportunities but contact between this group and the indigenous was minimal and deliberately restricted. This was the result of the former’s wish to restrict the latter’s access to power. Education in English was provided by missionaries and was in general a minimal existence.

After World War II Kenya aimed for independence and following the political trend at the time, the British relented. According to Schneider, “[i]n a sharp turn of their policy, [the colonisers’] goal now was to “modernize” these countries and prepare them for independence (amongst other things by teaching English on a broader scale), thus

building the ties that would preserve privileged cultural, political, and economic relationships in the future, in the form of membership in the Commonwealth” (2007: 192).

Accordingly, English was taught and learned by large numbers of people after Kenya’s independence in 1963. However Schneider contends that nativisation began in the late 1940s and its use only spread more widely after independence. KenE has had to compete with Kiswahili, a regional language frequently used for its unifying effects among various ethnic groups, in certain “high-language” domains (Schmied, 2004: 947). However Schneider, echoing the claim of Abdulaziz (1991) , asserts that the presence of KenE is “strong” in the country. He elaborates:

Certainly it has an essentially elitist and utilitarian character, as a language that is an indicator of a good education and the entry gate to desirable professions and white-collar jobs. Its use or non-use in any given situation depends upon a complex array of factors: social setting of the situation, medium, topic of conversation, status, ethnicity and language skills of interactants, and location. Its use is tied with upper- or middle- class status and with urban rather than rural contexts, but no longer exclusively so[.]
(2007:193)

Endonormative tendencies too are emerging (Kioko and Muthwii, 2004), but given that nativisation itself is ongoing and the country’s current social dynamics that do not favour KenE being used as a marker of national identity (Schneider, 2007:194), it is uncertain if the variety will be institutionalised in the near future.

3.3.8 Philippine English

Unlike most other varieties of English found throughout the world, Philippine English (PhilE) was derived from American English, both a colonial variety itself (Schneider, 2007: 251ff) and a principal norm provider of English (Romaine, 1998: 6). The history of PhilE began with the American colonisation of the islands in 1898, after more than three centuries of a previous colonisation by the Spaniards. Unlike the British who preferred a policy of “indirect rule” which delegated administrative powers to selected locals and thus restricting access to English for the populace (Fennell, 2007:193), the Americans opted for a more open approach on the basis of the perceived superiority of their culture and language that should be imparted to the colonised (Schneider, 2007:140). Nativisation began about a decade before independence in 1946. It was a process shaped by resistance to Japanese occupation around World War II and the lack of development of the national language (for which Tagalog/Filipino was designated). During the Marcos era in the 1970s, English was actively promoted and very highly regarded by the government (Schneider, 2007:141).

The widespread use of English continues to this day although “[n]owadays, English is very much the language of formal domains: business, politics, higher education, and parts of the media” (ibid.). PhilE is thus noted to be “monostylistic and close to written norms” (Gonzalez, 2004 in Schneider, ibid.). Nativisation in terms of vocabulary is also characterised by not only indigenous words but also Spanish loanwords. Of late, resistance to English as a marker of prestige has manifested itself although signs of endonormativity preferences have been detected. Interestingly a form of English,

called Taglish, has survived by mixing itself with Tagalog and serves as a marker of identity especially for city dwellers. Some signs of endonormative preferences have been identified although further development into institutionalisation is harder to say.

3.4 Concluding remarks

This chapter describes the other important component of the study. This is the process of nativisation. It is seen as part of the language change process and as such, working mechanisms identified both for language change and variation are believed to have the ability to explain the patterns of article usage observed in the varieties in later chapters.

The account of nativisation has also included a discussion of world Englishes, a framework that recognises many socio-political threads running through the phenomenon called nativisation. It introduced the varieties under study in their socio-historic contexts, an attempt to both chronologically contextualise these varieties and prove that the construct of ‘nativisation’ is linguistically tenable. The overview of nativisation in this chapter will help us to understand the various variation patterns of articles found in the Outer Circle which will be discussed in the chapters ahead.

Chapter 4 Research Design

This chapter describes the research tools and methods used in this study. It begins with a discussion of the corpus-based approach in 4.1. In 4.2 the components of the International Corpus of English, the Santa Barbara Corpus of Spoken American English and parts of the Frown corpus that form the database are described. It also includes a description of two other corpora which are used to inform the study. Finally the three stages of the analysis are described in 4.3.

4.1 Why a corpus-based approach?

Corpus-based studies of grammar can complement traditional descriptions available in grammar books as well as those that are based on intuition. Often these analyses lead to discoveries of mismatches between these descriptions and the actual patterns of use in the real world (McEnery and Wilson, 1996). In some cases, the new findings may challenge certain perceptions of grammar that have long been held by both grammarians and users (Meyer, 2002). This is elaborated by Collins:

Using a corpus, the linguist can make more objective statements than introspective analysis permits, and state frequencies in a variety of uses of the language. No individual can be expected to have an adequate command of the whole grammatical repertoire of a language, and no grammarian can be expected to describe adequately the grammatical properties of the whole repertoire from his or her own unsupplemented resources. Even when linguists are working in those areas of the repertoire in which they are most likely to be thoroughly competent (for example informal conversation) it is difficult to guard against idiolectal bias, accidental omissions and distortions, in the absence of objective data.

(1991:11)

Additionally the description can be made qualitatively where it is “based on the patterns of structure and use” (Biber and Conrad, 2001:332).

In chapter 1 it was argued that L2 studies of articles are usually methodologically restricted where researchers primarily rely on elicited data in experimental settings. While findings based on such data are important, corpora could offer a language-in-use perspective (i.e. Saussure’s *parole*) to the study of articles.

4.2 The International Corpus of English (ICE)

The ICE project began in 1990 with the aim of developing comparable corpora for the study of varieties of English or Englishes (Greenbaum, 1991). To ensure comparability, the same sampling frame was used for each variety. (The umbrella term ‘ICE’ explains why each corpus used in this study, except the American one, is referred to by some corpus linguists as a ‘sub-corpus’ or a ‘component’ of ICE.) The terms ‘sub-corpus’ and ‘component’ are used in the rest of this chapter for consistency. In subsequent chapters, only ‘component’ will be used and it will be interchangeable with ‘corpus’, e.g. ‘British component’ to ‘British corpus’ and vice versa.

Each sub-corpus is constructed from 500 samples containing 2,000 words each. Three hundred of these samples contain spoken language while the remaining 200 comprise written language. The sub-corpora therefore contain 1,000,000 words each. Each component representing one variety (with more currently being developed) was

compiled by a team of linguists usually based in the country where the variety is spoken. The text categories constituting the samples are described in table 4.1.

Table 4.1: ICE text categories (N.B: Number of 2,000-word files indicated in brackets.)

Spoken (300)	Dialogues (180)	Private (100)	Conversations (90) Phone calls (10)
		Public (80)	Class Lessons (20) Broadcast Discussions (20) Broadcast Interviews (10) Parliamentary Debates (10) Cross-examinations (10) Business Transactions (10)
	Monologues (120)	Unscripted (70)	Commentaries (20) Unscripted Speeches (30) Demonstrations (10) Legal Presentations (10)
		Scripted (50)	Broadcast News (20) Broadcast Talks (20) Non-broadcast Talks (10)
Written (200)	Non-printed (50)	Student Writing (20)	Student Essays (10) Exam Scripts (10)
		Letters (30)	Social Letters (15) Business Letters (15)
	Printed (150)	Academic (40)	Humanities (10) Social Sciences (10) Natural Sciences (10) Technology (10)
		Popular (40)	Humanities (10) Social Sciences (10) Natural Sciences (10) Technology (10)
		Reportage (20)	Press reports (20)
		Instructional (20)	Administrative Writing (10) Skills/hobbies (10)
		Persuasive (10)	Editorials (10)
		Creative (20)	Novels (20)

The texts in the corpus were collected from 1990 or later. Speakers or writers of the texts are at least 18 years old, and were educated in the English medium. Furthermore they were either born in the country where the variety is spoken, or moved there at an early age and were later educated in the English medium as spoken in that country.

Written and spoken samples were produced by both males and females. Although a wide range of age groups was sampled, the sampling proportions do not reflect the populations in the countries. In technical discourses, for example, the tendency of males to dominate the production of such language is not necessarily reflected in the sampling.

With the exception of the British, New Zealand and Australian components, all other ICE components used in this study were downloaded from the project's website. A compact disc containing ICE-GB was obtained from the coordinator of the project at University College London. The ICE-NZ data were made available from Victoria University of Wellington by its coordinator. ICE-Aus, held at Macquarie University, was accessed by the internet.

4.2.1 Santa Barbara Corpus of Spoken American English

The Santa Barbara Corpus of Spoken American English (SBC) is a collection of naturally occurring spoken interaction sampled in the United States. The corpus also forms part of ICE. The SBC provides the main source of data for the surreptitious spoken samples of the American component of ICE. A compact disc containing the part

one of SBC was obtained from the Linguistic Data Consortium, University of Pennsylvania. In this study, data from SBC (part one) were combined with some written data from the Freiburg-Brown Corpus of Written American English (Frown).

4.2.2 Freiburg-Brown Corpus of Written American English (Frown)

The Frown corpus has a million words of written American English. It was made to match the Brown corpus, which was built in the 1960s, using the same sampling frame but with a 30-year interval. Frown consists of data from the early 1990s. The Frown text types selected for this study are learned and scientific (category J), popular lore (category F), reportage (category A), skills, trade and hobbies (category E), editorial (category B) and general fiction (category K). These are to correspond to the ICE categories of academic writing, non-academic writing, reportage, instructional writing, persuasive writing and creative writing respectively. The corpus, held at Macquarie University, was accessed by the Internet.

4.2.3 Corpus of Contemporary American English and Brigham Young University – British National Corpus

Given that each ICE component has only a million words – a small number by today's corpus standards, many native usage patterns are not attested in the data. This deficiency does not augur well for the usage patterns found in Outer Circle varieties. While some of them may appear untypical or marked to the casual observer, it is possible for a number of them to be found in Inner Circle usage, albeit in lower frequencies. To ensure that any absence of similar usage patterns in the Inner Circle

corpora is not indiscriminately interpreted as proof for markedness (or worse, unacceptability), two large Inner Circle corpora were used for comparison in chapters six and seven. These are the Corpus of Contemporary American English (COCA) (Davies, 2009) and the Brigham Young University – British National Corpus (BYU-BNC) (Davies, 2004).

These two corpora are publicly available and accessible online. COCA contains more than 385 million words, with 20 million words each year from 1990-2008. Its categories include spoken, fiction, popular magazines, newspapers, and academic texts. The corpus is updated every six to nine months. BYU-BNC is modelled on the original BNC held at Birmingham University but was compiled by Davies (2004). Together these two corpora are seen as representative of the two major norm-providing varieties in the English-speaking world today. Rather than strictly being confined to the limits of ICE especially where native usage is concerned, the use of these additional data is seen as erring on the side of caution.

4.3 Levels of analysis

This study is exploratory. It contains three levels of analysis, each a corpus-based study in its own right.²⁵ The design of the overarching study is closely tied to the types of techniques which are commonly used in corpus linguistics. These techniques include frequency count, collocational analysis and annotation.

²⁵ Each level of analysis is treated as a full-length study in its respective chapter.

The first stage is a study that examines the frequency of overt articles in all eight corpora. The decision to designate frequency count as the first level of analysis was due to its perceived 'obviousness' and the fact that a previous study on the use of articles in world Englishes was based on it, i.e. Sand (2004).

The second level of analysis is a study of collocation. Due to the nature of this analysis, the focus is on overt articles again. Two types of analysis were performed: two-word level collocation and multiple-word level collocation. Subsequently, the collocates of these two-word or multiple clusters were analysed.

The third level is a study based on annotation. The taxonomies of usage types that form the annotation schemes are described in chapter seven. This study includes a case analysis of the word 'house' in which the patterns of \emptyset were investigated. A synthesis of the findings from each of the three studies is presented in chapter eight.

All frequency and collocational analyses were performed using Wordsmith Tools 4.0 (Scott, 2004). This concordancing software also performs statistical analyses such as log-likelihood and mutual information to assess the strength of the collocational patterns described in chapter six. The chi-squared statistic used in chapters five and seven was performed using the statistical program R.

Chapter 5 Frequency and Variation in Article Use

An observation that is commonly made of article variation is that it involves deletion and insertion of articles. This chapter presents a study that quantitatively investigates this claim. The next section briefly discusses the relevant findings of previous studies in both world Englishes and SLA. 5.2 and 5.3 describe the analysis presented in this chapter. Sections 5.4 through to 5.8 present the findings of an examination of the frequencies of definite and indefinite articles in the varieties and discuss what they mean in terms of nativisation.

5.1 Article deletion and insertion

Descriptions of the varieties offered in a handbook of varieties of English (Kortmann et al., volume 2, 2004) make it clear that the irregular use of articles often boils down to deletion and insertion of articles. Deletion appears to be more common than insertion. For instance, in a description of Singaporean English, Wee (2004:1061) highlights the absence of articles in the NP structure and quotes Platt and Weber, who state that “[i]t is noticeable that a definite or definite article does not always occur in [Colloquial Singapore English] in positions where it is obligatory in [Standard British English]” (1980:70 in *ibid.*) Wee illustrates the omission of an article with the following sentences (*ibid.*):

58. I don't have *∅ ticket*.
59. Maybe you better have *∅ microphone ∅ bit* closer.
60. She got *∅ car* or not?

61. She buy \emptyset dress for what?

Describing East African English, Schmied similarly points out that “[a]rticles ... tend to be omitted” (2004:932). He suggests that the omission is “partly ... an overgeneralization of British usage, e.g. *I am going to church/school/*post office*” (ibid.). Similarly, in summarising the features that are found in the Englishes of Africa, South and Southeast Asia, Mesthrie notes that omission of articles is pervasive in these varieties (2004:1138). Burridge (2004: 1148), too, highlights the prevalence of deletion in Aboriginal English in Australia.

Deletion of articles can also be found in Inner Circle varieties. Algeo lists, among others, “all afternoon/morning/evening” in American English. These usages stand in contrast to “all the afternoon/morning/evening” in present-day British English (2006:43).

Insertion of articles, however, is less numerous. Some examples can be found in Sand (2004:291):

62. The most universalized celebration is the Christmas <ICE-Ken: exam essay>.

63. We don’t need to go to the university to acquire you know a degree in mathematics simple arithmetic will get you the answer <ICE-SIN S2A-002>.

64. We present a new concept in the broadcasting in India <ICE-IND S2A-007>.

As for Irish English, which is a native variety for many of its speakers, Filppula notes that “[f]rom very early on ..., writers on [Irish English] have noted the Irish predilection

for using the definite article in contexts where it is not used in [Standard English] (1999:56). Some examples from Hickey (2007:251) are listed below:

65. *The youth* now isn't inclined to take on anything.
66. God, I'm parched with *the thirst*.

It has to be admitted that most cases of deletion and insertion involve the definite article. Nevertheless, the indefinite article *a/an* is also found deleted and inserted as shown in (58), (59), (60) and (61) above and in these examples:

67. The impact of MDT on the prevalence and incidence rate of leprosy was studied in a project area of 1,300,000 population over a period of 14 years, with *a special reference* to new monolesionol cases and those presenting as smear positive cases ("Impact of MDT on leprosy with a special reference to monolesions and smear positivity among the new cases" by Nanda Kishore in *Indian Journal of Dermatology, Venereology and Leprology*, (2002:133, volume 68, number 3).
68. Ok, you want to go to watch Ø movie or you want to go to.... (from Wee and Ansaldo, 2004:59).

In SLA, one of the causes of learners' frequent omission and misuse of articles is said to be L1 transfer (e.g. Ionin et al., 2008; Garcia Mayo, 2008; Thomas, 1989). For example, Garcia Mayo found that her Spanish-speaking participants transferred their L1 knowledge of articles to English, resulting in overuse of *the* in what she calls "cultural, structural and general reference" usage types (see Liu and Gleason's categories in figure 7.1). Similarly, Ionin et al. (2008) found that their Spanish speaking participants

performed better in supplying the correct articles in their experiment than their Russian speaking ones. Unlike Spanish, Russian has no articles. L1 background thus seems to be an appropriate variable to include in the present study. However, one drawback of the data from ICE is that information concerning speakers L1s is mostly unavailable (although other ‘non-world Englishes’ corpora such as ICLE ²⁶ may include such information). In view of the lack of this information, we will provide a broad profile of definiteness marking in all the L1s that are most commonly used by the speakers in each Outer Circle variety. For India, these include Hindi/Urdu, Gujarati, Punjabi, Tamil and Kannada. For Singapore, these are Mandarin, Hokkien, Malay and Tamil. For Kenya and the Philippines, they are Kiswahili and Tagalog respectively. These languages are claimed to be the major ones in their respective countries (Kortmann et al., 2004). Table 5.1 presents the profiles of these languages in relation to the presence/absence of articles and the Outer Circle varieties they are found to co-exist with.

Based on table 5.1, it would be interesting to see how much PhilE speakers’ use of articles parallels with native norms when compared to the other varieties.

²⁶ International Corpus of Learner English (Granger, 1998).

Table 5.1 Profiles of L1s found with IndE, SgE, KenE and PhilE

	[+Articles]	[-Articles]
IndE	-	All
SgE	-	All
KenE	-	Kiswahili
PhilE	Tagalog (definite <i>ang</i> , indefinite <i>isang</i>)	-

SLA research has also highlighted the significance of learners' proficiency in accounting for article deletion and insertion. Liu and Gleason, for example, discovered that overuse of *the* increased as their Chinese speaking participants' proficiency "improved from low level to intermediate level ... and then began to decrease as their English proficiency improved further" (2002: 16). To compare Liu and Gleason's finding which was based on article-less L1s, Garcia Mayo examined Spanish-speaking participants and found that "accuracy of article use increased with proficiency" (2008: 560). Although L1 transfer is again shown to have some effects, proficiency also appears to be an important variable that interacts with it.

As with L1s, ICE does not include much information about (non-native) speakers' proficiency. Even in the category of non-professional writing, comprising students' untimed and examination essays, this information is not indicated. Nevertheless we know that the "texts originate from adult speakers (over 18) who received their formal education through the medium of English to ensure a certain standard of linguistic

skills” (Hudson-Eattle and Schmied, 1999: 5). Therefore, while we can assume that there are different levels of proficiency for the speakers of Outer Circle varieties in the data, none of them can be regarded as elementary. Furthermore, we can also infer that instances of less proficient language are more commonly found in spoken than written data. To the extent that these variables can be examined through the corpus data, we will see how L1 transfer and proficiency influence article use in Outer Circle varieties.

In summary, if article deletion and insertion are as widespread as they are claimed to be by these authors, one should be able to find clear differences in the frequencies of articles in the different varieties, especially in the Outer Circle ones. This possibility can be easily transformed into a corpus-based research question. On the basis of the findings, it can be ascertained whether overuse and underuse of articles are genuine patterns of variation. The first stage of the study (see chapter four) was designed to answer this question.

5.2 The study

The study compares the frequencies of the definite and indefinite articles that are obtained in the spoken and written text types in the corpora. In simpler terms, it sets out to investigate the overuse and underuse of articles. A similar study was conducted by Sand (2004) in which she investigated the frequencies of the definite and indefinite articles. One of her findings will be discussed later in the chapter. A major difference between her study and the present one is the selection of the varieties that form the dataset. Whereas Sand’s study examined ICE-GB, ICE-NZ, ICE-Jamaica, ICE-Sin,

ICE-Ind, ICE-Kenya, the German component of the International Corpus of Learner English (ICLE-German) and a spoken corpus representing Irish English (Northern Ireland Transcribed Corpus of Speech/NITCS), this study excludes ICE-Jamaica, ICLE-German and NICTS but includes ICE-AUS, ICE-Philippines and part 1 of the Santa Barbara Corpus as well as parts of FROWN to represent American English. The exclusion of ICE-Jamaica and NICTS is due to their inaccessibility to the researcher. The other learner corpora are not included due to their largely interlanguage nature. The final selection of the eight corpora in this study is also in part due to the desire to provide a balanced comparison between four Inner Circle varieties and four Outer Circle varieties.

A second major difference between the two studies concerns the statistical procedure each employs to interpret its findings. Sand's study utilised an index of tokens per thousand words to normalise the frequencies across the differently proportioned text types. All comparisons were then made on this basis. While this study makes use of the same index to achieve normalisation, it is further reinterpreted as the actual frequency of an article obtained in a thousand words of a given text type. This in turn allows a chi-squared test to be performed and, by means of statistical significance, attains greater confidence in stating observations that are based on the different frequencies.

5.3 Procedures

As described above, a normalised figure in this study is reinterpreted as an actual occurrence in a random thousand words in a variety and not as a degree of frequency.

To put this into effect, all normalised figures are rounded to the nearest whole integer. This also means that the indices of token-per-thousand-words are transformed into nominal data and lose their function as ordinal data. This reinterpretation subsequently allows the chi-squared test to be performed to determine if the differences in the frequencies of an article in the reconfigured sample of a thousand words from each variety are statistically significant.²⁷

The following frequency comparisons were made:

- 1) comparing frequencies obtained from each text type *within* a variety
- 2) comparing frequencies obtained from each text type *across* varieties
- 3) comparing frequencies obtained from the text types in Outer Circle varieties with those from the text types in Inner Circle varieties.

Differences arising from the comparisons in (1), (2) and (3) were subjected to a chi-squared test.

Where necessary, KenE and AmE were excluded from the tests. This is because of the frequencies that were unobtainable from their data, making these two varieties incomparable with the others where comprehensiveness was concerned. There are two reasons for this incomparability. First, unscripted monologue is not available in ICE-Ken (see Hudson-Ettle and Schmied, 1999). Second, for AmE public dialogue,

²⁷ While statistical significance is important in stating observations based on frequency distributions (or for any quantitative study), it is equally important to remember that words are not used in random order. Thus sample size does influence proportions. These two facts must be considered in any quantitative investigation of textual data (Baayen, 2008:7-8). However, because the sampling frame of ICE is identical for all sub-corpora and because the texts of SBC and FROWN have been made to fulfill the same sampling requirements of ICE as much as possible, the reconfigured sample of 1,000 words from each variety are, in a sense, truly random.

unscripted monologue, scripted monologue, nonprofessional writing and correspondence were not available in the texts selected for that variety (see 4.2.1 and 4.2.2). All repeated tokens were eliminated (e.g. *the, the, the man* was counted as one token). Other irrelevant tokens were similarly excluded (e.g. *example (a)* was not counted).

5.4 Results: *the*

This section describes the results of the frequency analysis. They will be discussed in the next section. Table 5.2 shows the frequencies of the definite article per 1,000 words.

Results for comparison (1) are presented first. The chi-squared test results indicate that the frequencies obtained in each variety are statistically significant.²⁸ This means that the occurrence of at least one text type in any given variety was found to be higher or lower than what would be obtained by chance. To ascertain which of these frequencies have contributed significantly to the chi-square values, a post hoc analysis was conducted for each of the varieties to identify the difference (or residual) of the observed frequency and the expected frequency.

²⁸ The chi-square values for the varieties are as follows: Singapore (χ_2 (11) = 55.5258, $p < 0.05$), India (χ_2 (11) = 58.3614, $p < 0.05$), Philippines (χ_2 (11) = 42.7737, $p < 0.05$), Kenya (χ_2 (10) = 30.2634, $p < 0.05$), Britain (χ_2 (11) = 41.4519, $p < 0.05$), Australia (χ_2 (11) = 27.0976, $p < 0.05$), New Zealand (χ_2 (11) = 80.5546, $p < 0.05$), and United States (χ_2 (6) = 21.1914, $p < 0.05$).

Table 5.2 Frequencies of *the* across varieties and text types

	S1A	S1B	S2A	S2B	W1A	W1B	W2A	W2B	W2C	W2D	W2E	W2F	Average
SIN	32	59	58	78	88	41	87	62	64	65	88	54	65
IND	31	70	69	92	63	46	74	78	82	79	105	66	71
PHI	35	64	67	86	84	54	81	80	91	71	94	68	73
KEN	36	55	NA	65	68	39	75	66	74	52	70	53	59
GB	29	54	70	74	76	46	79	77	70	70	79	55	65
AUS	32	51	62	69	77	56	70	68	71	72	72	58	63
NZ	33	58	69	74	87	64	121	110	73	93	71	99	79
US	82	NA	NA	NA	NA	NA	33	66	60	60	61	56	60
Average	39	59	66	77	78	49	78	76	73	70	80	64	

Table 5.3 indicates the residual values, standardised to z-scores, as obtained from the post hoc analysis. Using an alpha value of 0.05, the critical value for a two-tailed test is ± 1.96 (i.e. mean-to-z is 0.475). Therefore, a positive value higher than 1.96 indicates an overrepresentation (i.e. overuse) of the definite article in that text type. Conversely, a negative value lower than -1.96 indicates an underrepresentation (i.e. underuse) of *the*. Such residuals are boldfaced in table 5.3. These are the significant contributors to the chi-square relationship between a variety and the text types that represent it.

It can be seen from the table that in SgE the definite article is overrepresented in nonprofessional writing (W1A), academic writing (W2A) and persuasive writing (W2E). In IndE it is overrepresented in scripted monologue (S2B) and persuasive writing (W2E). As for PhilE, *the* was found to occur significantly more than by chance in reportage (W2C) and persuasive writing. KenE has *the* overrepresented in only one of its categories, i.e. academic writing. As for two of the Inner Circle varieties, BrE and AusE, *the* is not overrepresented in any of their categories. The other two Inner Circle varieties, however, tell a different story. The definite article was found to be overrepresented in academic writing, non-academic writing and creative writing in NZE. AmE, on the other hand, has *the* overrepresented only in private dialogues.

Table 5.3 also shows that the definite article is underrepresented in private dialogue in all varieties except for AmE (where, in stark contrast, it is overrepresented). Other than private dialogues all Outer Circle varieties (i.e. SgE, IndE, PhilE and KenE) and BrE have *the* underrepresented in correspondence. AusE does not have any other category

in which *the* is underrepresented. On the other hand, NZE has public dialogue (S1B) and AmE has academic writing with *the* underrepresented in them in addition to private dialogue.

Table 5.3 : Residuals in a chi-square relationship between a variety and its text types for *the*

	S1A	S1B	S2A	S2B	W1A	W1B	W2A	W2B	W2C	W2D	W2E	W2F
SIN	-4.06223069	-0.70467267	-0.82902667	1.65805334	2.90159335	-2.94304469	2.77723935	-0.33161067	-0.08290267	0.04145133	2.90159335	-1.32644268
IND	-4.7684085	-0.1480872	-0.2665570	2.4582478	-0.9773756	-2.9913618	0.3257919	0.7996710	1.2735501	0.9181408	3.9983549	-0.6219663
PHI	-4.440345	-1.044213	-0.692889	1.532163	1.297947	-2.215293	0.946623	0.829515	2.117703	-0.224457	2.469027	-0.575781
KEN	-3.0323560	-0.5663544	NA	0.7315411	1.1209098	-2.6429874	2.0294367	0.8613307	1.8996472	-0.9557231	1.3804889	-0.8259335
GB	-4.4577727	-1.3549147	0.6309145	1.1273717	1.3756004	-2.3478292	1.7479433	1.4997147	0.6309145	0.6309145	1.7479433	-1.2308003
AUS	-3.9214472	-1.5308323	-0.1467921	0.7339607	1.7405354	-0.9017232	0.8597826	0.6081389	1.1114262	0.9856044	1.1114262	-0.6500795
NZ	-5.2019443	-2.3951398	-1.1601458	-0.5987850	0.8607534	-1.7215067	4.6780074	3.4430135	-0.7110571	1.5343864	-0.9356015	2.2080195
US	2.88394808	NA	NA	NA	NA	NA	-3.45704033	0.81342125	0.03697369	0.03697369	0.16638162	-0.48065801

The chi-squared test results for comparison (2) are presented next. It was found that the frequencies obtained in all the varieties for the categories of public dialogue, unscripted monologue, scripted monologue, nonprofessional writing, correspondence and reportage were not statistically significant.²⁹ On the other hand, the differences in the frequencies obtained in all the varieties for the categories of private dialogue, academic writing, non-academic writing, instructional writing and creative writing are significant.³⁰ Table 5.4 shows the residuals, converted to z-scores, for chi-squared values that are statistically significant. Using an alpha of 0.05, values higher than 1.96 and lower than -1.96 are boldfaced.

For private dialogue, AmE emerges as the sole contributor to the statistically significant chi-squared value. For academic writing, NZE's unusually high frequency and AmE's low occurrence contributed to the score. As for non-academic writing, NZE's high use of the definite article was found to affect the significance score. In instructional writing, KenE's low frequency and NZE's high frequency contributed to the score. IndE's high occurrence of *the* and AmE's low use of it jointly contributed to the significant chi-squared value of persuasive writing. Finally, NZE again solely affected the significance score in creative writing with its high frequency.

²⁹ The chi-squared values are as follows: public dialogue ($\chi_2(6) = 4.2822$, $p > 0.05$), unscripted monologue ($\chi_2(5) = 1.7443$, $p > 0.05$), scripted monologue ($\chi_2(6) = 6.9331$, $p > 0.05$), nonprofessional writing ($\chi_2(6) = 7.035$, $p > 0.05$), correspondence ($\chi_2(6) = 9.7052$, $p > 0.05$), and reportage ($\chi_2(7) = 9.147$, $p > 0.05$).

³⁰ The chi-squared values are as follows: private dialogue ($\chi_2(7) = 56.0387$, $p < 0.05$), academic writing ($\chi_2(7) = 52.2839$, $p < 0.05$), nonacademic writing ($\chi_2(7) = 21.5733$, $p < 0.05$), instructional writing ($\chi_2(7) = 15.1388$, $p < 0.05$), persuasive writing ($\chi_2(7) = 18.65$, $p < 0.05$) and creative writing ($\chi_2(7) = 25.8684$, $p < 0.05$).

Table 5.4 Residuals in a chi-square relationship between a text type and varieties for *the*

SgE	IndE	PhilE	KenE	BrE	AusE	NZE	AmE	
S1A	-1.0843461	-1.2449900	-0.6024145	-0.4417706	-1.5662777	-1.0843461	-0.9237022	6.9478472
W2A	1.0791275	-0.3975733	0.3975733	-0.2839809	0.1703886	-0.8519428	4.9412680	-5.0548603
W2B	-1.5928821	0.2439549	0.4735596	-1.1336729	0.1291526	-0.9040682	3.9176290	-1.1336729
W2D	-0.6263775	1.0439625	0.0894825	-2.1774074	-0.0298275	0.2087925	2.7143024	-1.2229275
W2E	0.8944272	2.7950850	1.5652476	-1.1180340	-0.1118034	-0.8944272	-1.0062306	-2.1242646
W2F	-1.2066653	0.2977486	0.5484842	-1.3320332	-1.0812975	-0.7051940	4.4348869	-0.9559297

Comparison (3) required that the varieties be divided into Inner Circle and Outer Circle groups. To ensure total comparability KenE and AmE were not included (see 5.3). The Inner Circle group comprises BrE, AusE and NZE while the Outer Circle group consists of SgE, IndE and PhilE. The frequencies obtained from these two groups are shown in table 5.5.

Table 5.5 Frequencies of *the* in Inner vs. Outer Circle varieties

	S1A	S1B	S2A	S2B	W1A	W1B	W2A	W2B	W2C	W2D	W2E	W2F
In. Circle	98	193	194	256	235	141	242	220	237	215	287	188
Out. Circle	94	163	201	217	240	166	270	255	214	235	222	212

When compared, the overall difference of frequencies obtained in the Inner Circle group and the Outer Circle group was found to be significant ($\chi^2(9) = 23.8946$, $p < 0.05$). A post hoc analysis indicates that the source of this significant value is located in persuasive writing (W2E). With an alpha value of 0.05, the residual is slight, but in different directions. In this text type the definite article was found to be underrepresented in Inner Circle varieties (-1.986316) and overrepresented in Outer Circle varieties (1.979567).

5.5 Discussion: *the*

What do the frequencies tell us about variation of the use of the definite article in the varieties? In 5.1 we considered the hypothesis that if a speaker's native language is article-less, the variety of English that s/he speaks will show a corresponding tendency

to omit articles. (This of course includes languages which mark definiteness by means other than articles, e.g. word order and stress placement.) If substrate influence is instrumental in making speakers delete and insert articles, we would then find that speakers of SgE, IndE and KenE to demonstrate these predilections in their respective corpora because their substrate languages do not have articles (see figure 5.1). In particular speakers of these three varieties would delete more definite article than Inner Circle speakers. PhilE speakers, on the other hand, would not delete *the* as much because Tagalog does have a definite article.

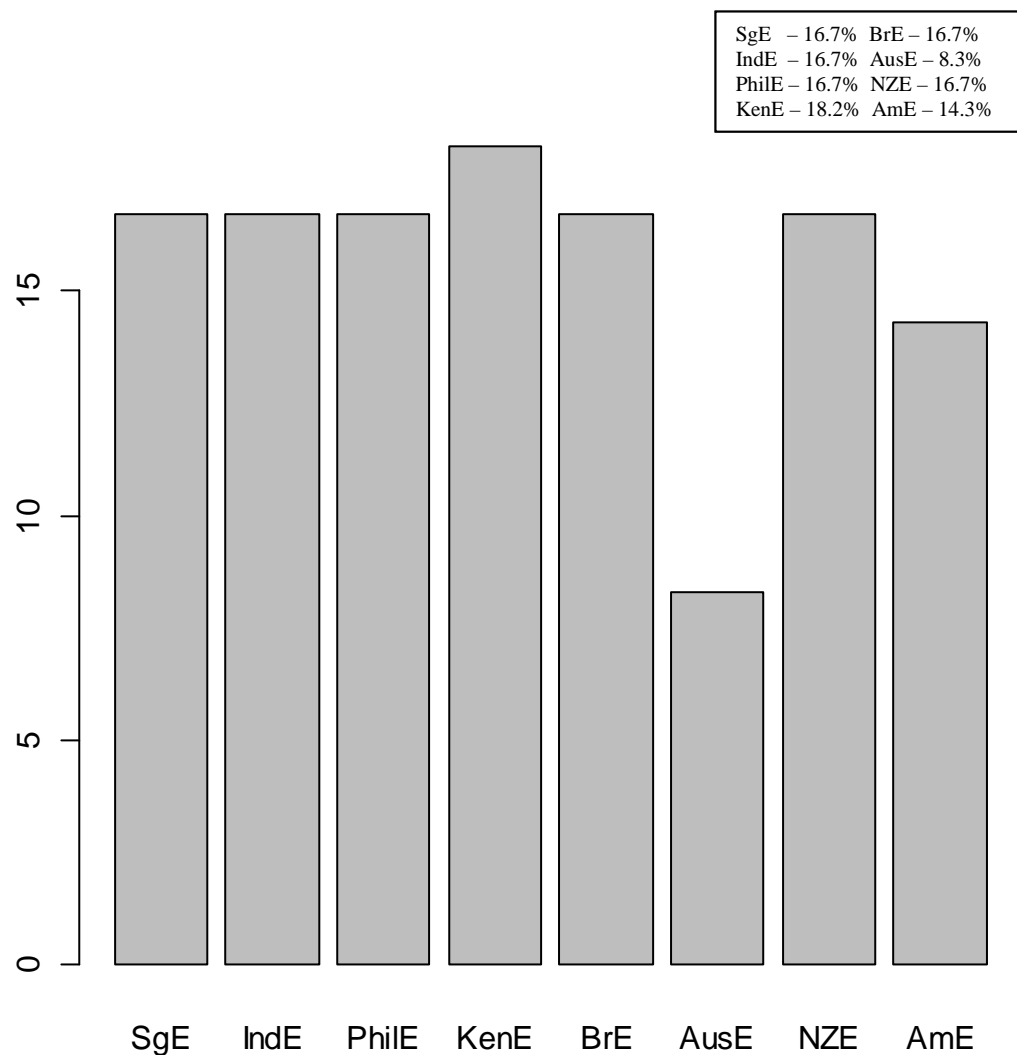
In comparison (1), it was shown that the definite article can be overused and underused in a single variety. If we assume that *the* is used equally frequently in all text types, what the results show is that in each variety there is an inherent variation in terms of how often the article is used. In SgE, for example, *the* is overused in nonprofessional writing, academic writing and persuasive writing but underused in private dialogue and correspondence. On the other hand, the situation is almost reversed in AmE where the definite article is overused in private dialogue but underused in academic writing. But what do these differences mean in terms of variation?

It is important for us to put the differences from comparison (1) into perspective. If one were a non-speaker of any of the varieties (but understood English), reading and listening to these Englishes would be akin to observing the results in table 5.3. Thus, from this viewpoint, the majority of speakers of SgE, IndE, PhilE, KenE, BrE, AusE and NZE were found to omit *the* significantly more frequently in private dialogue than

elsewhere *in their own varieties*. This is not the case of AmE speakers. From this perspective too, the closest case to homogeneity among the varieties would appear to be private dialogue, followed by correspondence. We will examine this finding further through the results of comparison (2) later. In terms of overuse, the category of persuasive writing has the most number of varieties overusing *the*. This place is shared with academic writing although AmE is also found to significantly underuse *the* in it. In comparison (2), we will evaluate further the validity of this finding.

Regardless of the exact categories in which *the* was overused or underused, we can evaluate the L1 transfer hypothesis using the combined totals of categories where there is incidence of overuse or underuse of *the*. The percentages of the underuse and overuse of the definite article in the varieties are depicted in figures 5.1 and 5.2 respectively.

Figure 5.1: Percentages of underuse of *the* in the varieties

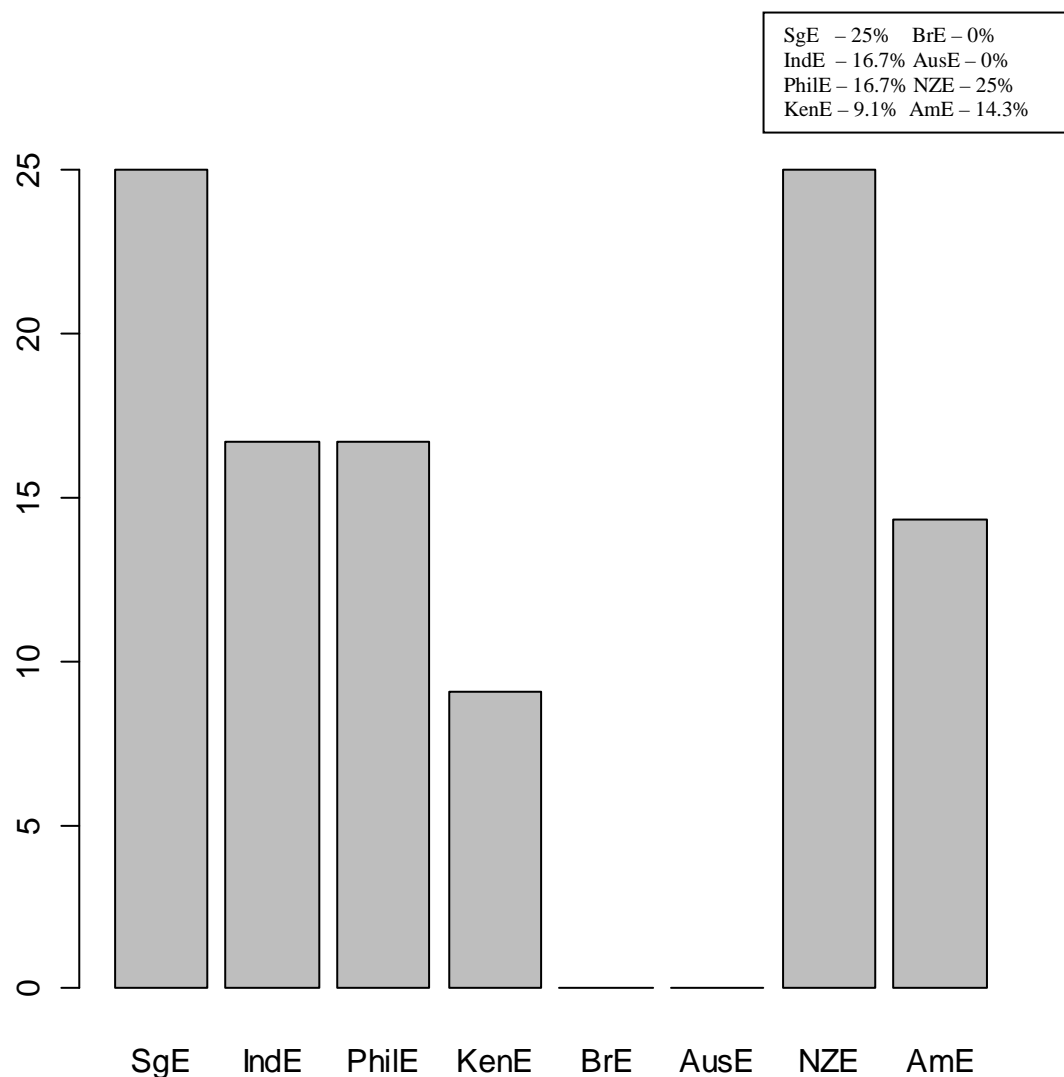


The bar chart in the figure above shows that the Outer Circle varieties are more uniform in their levels of the underuse of *the*. Thus, although an outside observer would notice variable levels of *the*-omission in the categories of each variety, there is an overall uniformity among many of the varieties in regard to how much omission was made overall. And if we leave AmE and KenE out of this analysis due to their dissimilar

corpus designs, it is the Outer Circle varieties that are shown to be most regular. Despite the earlier speculation, PhilE speakers exhibited as much underuse as SgE and IndE ones. This is contrary to Sand's findings where Jamaican English and Northern Irish English, two varieties with "a substrate containing definite articles", were found to demonstrate "significantly higher frequencies than all other varieties" (2004: 288). In the Inner Circle, the trends are far less predictable. AusE speakers were found to omit the least amount of *the* in their variety. However, this also means that the other two Inner Circle varieties, BrE and NZE, omitted as many definite articles as their Outer Circle counterparts. According to these findings, we can safely conclude that the L1 transfer hypothesis is not borne out.

Figure 5.2 depicts the percentages of overuse of *the* in the varieties. The bar chart shows that NZE and SgE have the highest percentage of overuse. Two other Outer Circle varieties, IndE and PhilE, follow them at a rather distant second. Behind them is AmE followed by KenE. BrE and AusE are at the last place with no recorded overuse. To eliminate any possible distraction from AmE and KenE, the ranking was reconfigured without them and the result is an almost neat division between the two groups of varieties, save for NZE which occupies the top place with SgE. In Sand (2004), although this variety was made to look like it paralleled BrE in all aspects, the higher occurrence of *the* in NZE was not highlighted and it turns out that her argument begs the question. To account for this finding, we clearly need to differentiate the overuse and underuse that are claimed to be found in the Outer Circle varieties from those in the Inner Circle only.

Figure 5.2: Percentages of overuse of *the* in the varieties



Recall from the above discussion that all varieties except AmE were found to underuse *the* in private dialogue. Why is this so? One possible reason is that the definite article is regularly dropped in conversations due to increased familiarity and shared discourse context between speakers. Biber et al. point out that “[s]peakers tend to use the minimum description that they think will achieve successful reference from the hearer’s point of view” (1999: 233). According to them one way this reduced description can be

made is through the use of imprecise reference. An example is the use of *and stuff like that* in the following:

60. I'm I'm, how I'm not too sure how much they're going to check up the sort of figures on what you're getting in your grant *and stuff like that* <ICE-GB:S1A-079 #154>.

Compare the vagueness in (60) and the clarity found in a similar discussion of grants in instructional writing:

61. If they do decide to give you a discretionary grant for a designated course it will be for the same amount and conditions as if you were eligible for a mandatory grant; *but the amount and conditions of a discretionary grant for a non-designated course are determined by the LEA and may not be the same as a mandatory grant* <ICE-GB:W2D-003 #73>.

In adding further details to the description of the grant, the italicised clause has made use of three definite descriptions using *the*, i.e. *the amount and conditions*, *the LEA* and *the same (thing)*. Written language does appear to have the advantage over spoken language in this sense. Other ways of creating imprecision in conversations that are found in the data include the use of deictic words such as *this* and *that* as in:

62. you know i have a real writing talent and [,] and people could look at *that* and think oh this is a really clever woman and yet i'm so stupid <ICE- NZ:S1A-004#103:1-104>.

Compare (62) with (63) where, instead of 'that', a more precise and succinct co-referent *the structure...* is provided for the situation in which *[t]he baby gets sick, the centre closes down, the minder shifts away, the hours of work change, the fees go up,*

there's no after school care...:

63. Helen May Cook, a childcare activist, wrote in *Mind that Child*:
'During the early child-rearing years women who work lead a rather crazed existence, trying to manage by continually propping up a system that never stays stable for very long. The baby gets sick, the centre closes down, the minder shifts away, the hours of work change, the fees go up, there's no after school care...' She argues for 'a much broader look at the structure of the workplace in relation to the domestic world <ICE-NZ:W2B-018#187:1-189>.

In contrast to the use of vagueness, and thus the lower incidence of *the* in spoken language, modified NPs such as the in (63) above may have easily given rise to the use of *the* in written language.

Deictic *this* and *that* may also suffice in establishing non-textual reference. This is rare in written language but quite common in conversations. Thus the occurrence of the definite article may also be adversely affected. Consider:

64. A: Look at *this*
C: So they're pasted on uh
A: Not so huge uh this one
B: Because this is from imported from Indonesia [.]

<ICE-SIN:S1A-008#80:1 – 83>

However, it is not only vagueness in spoken language that may contribute to the reduction of *the*. Because of the low descriptive content of this word, it is not

uncommon to find it missing in conversations. An example is the use of ‘South Island’ in NZE where it occurs in with *the* 100% of the time in the written data when functioning as a placename. In contrast, the spoken data yield one token of *Ø South Island* in the following:

65. *Ø whole south island* and the north island um <.>th</.> er around of wellington i think porirua <ICE-NZ:S1B-002#29>.

Despite the evidence that has been marshalled above, data from AmE suggest none of it is true. It seems that in private dialogues, speakers of AmE significantly increase their use of the definite article. Confounding the matter further is the fact that the other Inner Circle varieties seem to parallel the Outer Circle ones in this category. Has AmE created a norm of its own, as it sometimes has for other grammatical items? This is difficult to prove at this stage. We will try to do so with the results of comparison (2).

With these findings, however, we can offer a tentative confirmation of the L2 proficiency hypothesis (see earlier, this chapter). If the assumption that less proficient speakers are mostly found in the private spoken category is correct, it appears that the Outer Circle varieties’ low frequencies of *the* in conversations and telephone calls reflect their lack of proficiency. However notice that they are not elementary-level learners and thus none of the frequencies is too low, as they would have been in an SLA study on beginners. Importantly, this does not mean other variables such as informality can be discounted. Informality does seem to be the more powerful factor, as we will see shortly (cf. Biber et al., 1999: 271).

The complexity involving unclear trends emerging from the frequencies obtained for private dialogue was found for correspondence too. In this category all the Outer Circle varieties were discovered to underuse the definite article while AusE and NZE did not. However, spoiling this otherwise perfect divide between Inner and Outer Circles is BrE, which patterned with the Outer Circle varieties. Correspondence, nevertheless, is to some extent similar to private conversation. There is a fair amount of informality contained in social letters that make up half of this category. Consider:

66. Looks like Wai Li and you are as like as chalk and cheese. It's good you're still good friends despite Ø difference <ICE-SIN:W1B-012#284:10 – 285>.

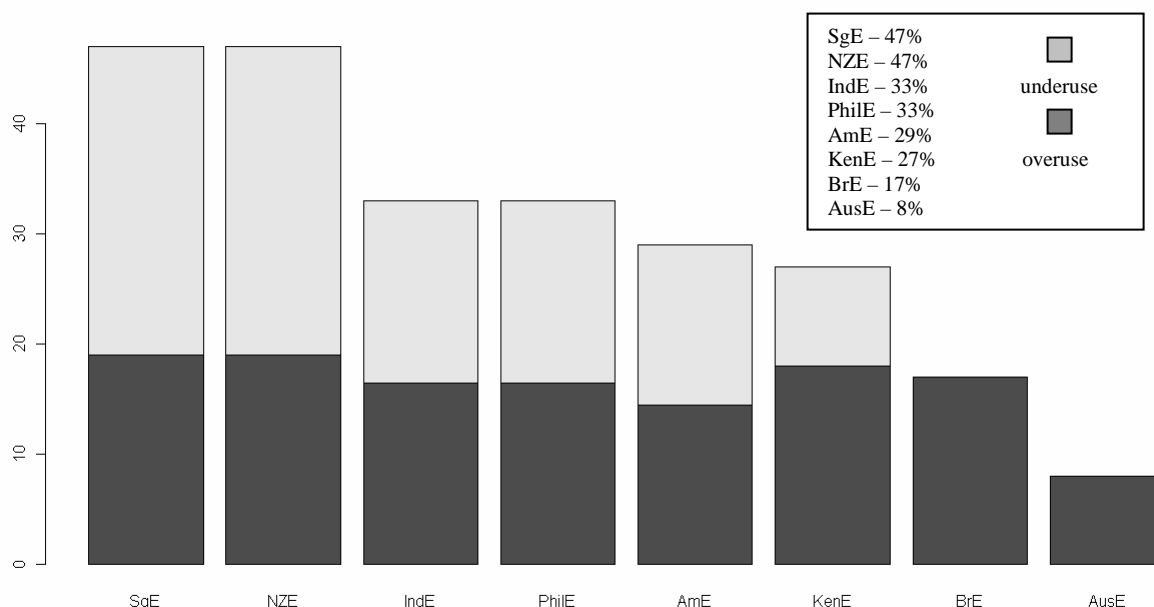
What is interesting, though, is the fact that all Outer Circle varieties underuse *the* in correspondence. It is possible that Outer Circle varieties' speakers tend to write social letters with the same level of informality that they use for speaking casually. But before we hasten to conclude that such is a feature of these varieties, recall that BrE was also found to do the same.

Any other observable trends are vague. Academic writing has two Outer Circle varieties overusing *the* (i.e. SgE and KenE) and with one Inner Circle variety doing the same (NZE) and another one doing the opposite (AmE). Public dialogue, scripted monologue, nonprofessional writing, non-academic writing, reportage and creative writing have either one Inner Circle or Outer Circle variety overusing or underusing *the*. Given that the definite article can be overused and underused in an unpredictable combination of categories for each variety, we can conclude at this juncture that we

need to be cautious about using these frequencies as an index of variation. This claim will be illustrated in the next analysis.

If overuse or underuse of the definite article by itself is not indicative of the variation that is thought to be inherent in the Outer Circle varieties, we will try to use *both* of them as an index of variation regardless of the positive/negative value. The varieties will then be ranked on this basis. Figure 5.3 is a bar chart showing the percentage of variable text types (as an index of variation) and the percentages for overuse and underuse for each variety.

Figure 5.3 Bar chart showing percentages of variation and proportions of overuse and underuse of the definite article in each variety



The above figure shows that the varieties with the highest level of variation are SgE and NZE. In the second place are IndE and PhilE. The third place is occupied by AmE, followed KenE in the fourth place. The fifth and sixth places are occupied by BrE and AusE respectively.

Even in this ranking it is difficult to see a clear divide between Inner Circle and Outer Circle varieties. NZE shares the top spot with SgE. In terms of underuse and overuse, it distances itself very distinctly from BrE, a variety that gave birth to it, and AusE, its closest cousin. This finding was not revealed in Sand (2004) because the differences between BrE and NZE were unified as ‘native’ patterns: “[i]f we look at the distribution in ICE-GB and ICE-NZ, we detect very similar distributions in the two

varieties....” (2004:287). Then when compared with the other varieties, the similarities and differences it shares with SgE (as discovered here) were masked.

AmE positions itself somewhere in the middle of this ranking. Nevertheless it must be remembered that in this study five text types were not available for this variety. Their inclusion in the AmE data could have revealed a different pattern. BrE and AusE are similar in their lack of overuse. Overall, three clear trends emerge from figure 5.1:

1. Most important of all, although Outer Circle varieties tend to delete and insert the definite article more frequently than Inner Circle varieties, NZE is a significant exception. Therefore, in general Inner and Outer Circle varieties do not differentiate themselves in terms of omission and deletion of articles in any clear, definitive way.
2. Variation that can be detected by frequency count is slightly more often due to omission than insertion. Of the 26 variable frequencies for all text types and varieties, 14 (54%) were underuse cases and 12 (46%) were overuse (see table 5.3). In addition all Outer Circle varieties, but not Inner Circle ones, were found to underuse the definite article. If this finding points to a variation pattern i.e. Outer Circle varieties underuse more definite article than Inner Circle varieties, it has to be cross-checked with comparison (2), which we will do below.
3. There is no firm correlation between article-less L1s and omission of articles by their speakers in English, thus discounting the L1 transfer hypothesis that has characterised much of SLA research on article acquisition. (As for the L2 proficiency hypothesis, only a weak support could be offered for it.) This is

evidenced by cases of overuse in the Outer Circle varieties, especially PhilE, and the fact that the Inner Circle varieties themselves underuse the definite article in a number of categories.

The discussion so far has been based on one assumption i.e. that the definite article is equally distributed *in all text types* in each variety. We saw that comparison (1) had characterised some of the text types with either an overuse or an underuse trend. However none of these was definitive and as a result, we need to address two issues that have come to our attention:

1. Why isn't there a neat division between Inner and Outer Circle varieties in terms of underuse and overuse of *the*? If we assume that the definite article is equally distributed *in all varieties* in each text type, what will the findings tell us?
2. Do Outer Circle varieties really omit more definite articles than Inner Circle varieties?

Comparison (2) is to examine the frequencies from the perspective of the text types. This is the perspective gained if one were to observe each text type produced by all speakers irrespective of their varieties. It is important to remember that all the frequencies obtained from comparison (1), e.g. lower frequencies of *the* in private dialogue compared to other categories, are immaterial in this analysis. We are in fact interested in how the production of each text type, say persuasive writing, is constrained by the varieties (see Tagliamonte, 2006:191-194 for a similar discussion).

Table 5.4 above indicates that some varieties were found to use the definite article more or less frequently than others in the categories of private dialogue, academic writing, non-academic writing, instructional writing, persuasive writing and creative writing.

For private dialogue, recall that AmE's unusually high frequency was in stark contrast to the relatively low, albeit uniform, scores in the other varieties. Based on the result of the post hoc test it is fair to conclude that AmE deviates from the norm expected for this category, effectively compromising any conclusion that points to the influence of face-to-face conversations and telephone conversations (private dialogue) on article omission. Closer inspection of AmE data reveals that there is a high number of repairs involving *the*. In some other varieties, especially Outer Circle ones, repeats are more common than repairs. Compare these two examples.

67. SAM: [There's] no problem. It will not take any frost. Soon as *the* b- --
ANGELA: Yeah [I-] --
SAM: [A little bit] of frost, ... it's gone.
ANGELA: I [,] I learned that. ... one time [*the*-] –

<SBC189.47 190.43 - 196.91 198.04>

68. Okay *the the the the* reason why I use the word communion is that it's important to relate to the audience <ICE-SIN:S2A-030#23>.

In (67) each of the two AmE speakers is found to use one repair involving *the*. As described in 5.3, repeats, but not repairs, were excluded from the analysis. This was to preserve the integrity of the dialogue. It is possible that repairs such as that in (67) have contributed to the anomalous frequency in AmE.

For academic writing, NZE's overuse and AmE's underuse of *the* again demonstrate the untenability of the L1 transfer hypothesis. It also demonstrates that certain text types do not necessarily have the kind of influence on article omission as previously thought. Biber et al., for example, claim that the definite article is "most common" in academic writing (1999: 267). It is notable that when varieties are made to function as independent variables, a text type can sometimes reveal differences even between two Inner Circle varieties.

The categories of instructional writing and persuasive writing have either one Inner Circle variety or one Outer Circle variety overusing or underusing the definite article (see table 5.4). Only SgE, PhilE, BrE and AusE produced all text types without unexpected levels of omission or insertion of the definite article.

The single most striking result to emerge from comparison (2) is NZE's propensity to overuse the definite article, as shown in its sole contribution for non-academic writing and creative writing as well as its role in the two categories indicated earlier, academic writing and instructional writing. A question that arose from comparison (1) is whether or not Outer Circle varieties underuse more definite articles than Inner Circle varieties (see above). We now see that this is not the case. Each variety has a variable distribution of *the* but when compared to one another, these distributional patterns do not necessarily coincide. However, from the perspective of the text types we now know that AmE has the highest level of underuse. This is in addition to overuse in private

dialogue. A question that needs to be asked is why NZE and AmE are so deviant from the norms of their related varieties.

We have seen that in AmE, the use of repair involving *the* (see 67 above) may have contributed to the unexpectedly high level of the definite article in private dialogue. In the case of NZE there is clearly more than one factor that contributes to the overuse in the four text types. Upon scrutiny, the NZE data were found to consist of several (i) collocations that are either absent or rare in other varieties (ii) collocations that are numerous but used in the same text and (iii) collocations that are formed with proper names, especially placenames. Figure 5.4 below is a comparison between NZE and BrE.

Figure 5.4: A comparison of selected collocates of *the* in NZE and BrE

		NZE	BrE
academic writing	the area	20	4
	the attenuation	16 (from 1 text)	0
	the availability	13 (11 from 1 text)	0
	the Bay of	15	0
	the Blue Mountains	11 (from 1 text)	0
	the British	34 (24 from 1 text)	3
	the concentration	15	0
	the New Zealand	33	0
	the North Island	16	0
	the South Island	21	0
	<i>the cow</i>	0	24(from 1 text)
non academic writing	the Auckland	42	0
	the Chatham(s)	28	0
	the community	11	4
	the company	25	7
	the North Island	19	0
	<i>the Scottish</i>	0	1
instructional	the Act	19 (from 1 text)	0

writing	the bank	32 (from 1 text)	0
	the boat	44	0
	<i>the ball</i>	<i>18 (from 1 text)</i>	<i>83(from2 texts)</i>
creative	the air	25	10
	the frontman	26 (from 1 text)	0
	the water	31	6
	<i>the midlands</i>	<i>0</i>	<i>2</i>

Note: unless indicated each frequency was obtained from at least 5 different texts; italicised items indicate a higher frequency in BrE.

The comparison above is not exhaustive and is only intended to be illustrative of the claims made regarding the overuse of *the* in the four text types in NZE. It is apparent from figure 5.4 that collocations that are formed with placenames are the most frequent ones in this variety. Interestingly, BrE does not make use of *the* with placenames or words deriving from them as much as NZE does. For example, *the Scottish* is found only once in non-academic writing in BrE. *The area* is an example of *the*-collocation that is far more frequently used in NZE than BrE. *The bank* illustrates the overuse of a specific *the* collocation in a single text, which tends to occur more frequently in NZE. This is not to say that such overuse does not occur in BrE. In fact, *the ball* and *the cow* are provided in the figure as counter examples. However the fact remains that such occurrences are far more common in NZE than in any other variety.

As for AmE, it is difficult to ascertain the factors that drove *the* to be underused in academic writing and persuasive writing. If we examine the top 100 most frequent words to collocate with *the* at the first position to its right (R-1) in AmE and, to

represent other ‘normal-range’ varieties, BrE, there is an important difference between them. Figure 5.5 shows this comparison in academic writing.

Figure 5.5: A comparison of the most frequent R-1 collocates of *the* in W2A in AmE and BrE

AmE			BrE		
	Word	Frequency		Word	Frequency
1.	first	8	1.	romans	10
2.	form	8	2.	roman	9
3.	other	8	3.	surface	9
4.	universe	8	4.	mine	7
5.	1970s	6	5.	process	7
6.	forms	6	6.	same	7
7.	initial	6	7.	filter	6
8.	observed	6	8.	first	6
9.	examples	5	9.	upper	6
10.	notebook	5	10.	Italians	5
11.	social	5	11.	more	5
12.	woman	5	12.	need	5
13.	work	5	13.	other	5
14.	document	4	14.	shaft	5
15.	last	4	15.	technological	5
16.	most	4	16.	business	4
17.	result	4			
18.	verbal	4			

This comparison is made using three randomly selected texts from the academic writing category in each variety. Notice that the top 100 most frequent R-1 collocates in academic writing in BrE consist of sixteen lemmas (e.g. *Romans* and *Roman* are counted separately). AmE yields eighteen. However, BrE is found to have many tokens of lemmas originating from single texts, e.g. *Romans* (10 tokens), *Roman* (9 tokens), *mine* (7 tokens) and *process* (7 tokens). *Surface* and *same* each come from two texts. In contrast, in AmE only *universe* and *notebook* are found in single texts. These facts by

themselves do *not* explain why AmE underuses *the* in academic writing but they do foreground the tightness of the topics found in BrE's academic writing. While this phenomenon has led to the overuse of *the* in NZE, the lack of this occurrence may help explain AmE's underuse. It is possible that (i) referents are often not 'picked up' and elaborated in the AmE academic prose data and (ii) *the* is frequently used non-referentially in this variety and are thus not referred to/described again.³¹ Consider (69a and b) for (i) and (70) for (ii).

69a. This suffix is not described by Mason in his discussion of *the verbal morphology*, nor in Turner, so that it appears to be quite spurious, and given the lack of documentation, one despairs of tracking it down <Frown: J35 152 – 154>.

69b. *The Romans* first conquered Italy, and then mainly with Italian arms the dominions overseas which ultimately extended from the Solway Firth to the Sahara and from Rhine and Danube, or even beyond, to Euphrates. The conquest of Italy was certainly not a process of enslavement. It was attacks by their neighbours, or the fear of such attacks, that gave early Rome the pretexts or motives for reducing them to submission; in addition, by confiscating some of their lands, *the Romans* were able to satisfy the land-hunger of their own peasantry <ICE-GB:W2A-001 #1:1>.

70. The Social Action Program of 1974 was *the result*. The program promised action "to achieve equality between men and women as regards access to employment and vocational training" <Frown: J22 81- 84>.

³¹ These observations are also true for persuasive writing.

In (69a) the NP *the verbal morphology* in the rheme of the first clause is not picked up in the following one. In (70) the same thing happened although this time it involves the non-referential NP *the result*. (69b) is shown as an example of anaphoric reference of the NP *the Romans* which is more commonly found in BrE than AmE.

To recapitulate, it was shown that definite article omission and insertion are not contingent upon

- (1) whether or not a variety is from the Inner or Outer Circle;
- (2) the L1 transfer hypothesis;
- (3) whether or not a text type influences the use of an article.

Proficiency may have a role to play in the frequencies obtained but the nature of the data does not allow this to be proven beyond doubt. Other than that, it was shown that NZE, an Inner Circle variety, had the highest amount of overuse. Further analysis pointed to this variety's tendency to use with *the* R-1 collocates that are formed from proper names, absent or rare in other varieties and originate copiously from a single text. Finally, it is AmE, and not the Outer Circle varieties, that tends to underuse *the*. The underuse is prevalent in academic and persuasive writing due to the less tight focus of the topics found in them. It is difficult to say if this pattern is due to specific styles of writing in AmE or simply a sampling difference between Frown and ICE.

The last comparison provides yet another perspective on the trends that have been observed. When the frequencies of the Inner and Outer Circle varieties are combined

accordingly, the chi-squared test result is significant (see 5.4). A post hoc analysis shows that the one text type in which the native and nativised groups differ significantly from each other is persuasive writing. The Inner Circle group shows a slight underuse while the Outer Circle group shows a slight overuse. It seems that the more dramatic differences between individual varieties observed in the previous two comparisons are masked with the combined totals. On the basis of these collective profiles, it can be said that Inner and Outer Circle varieties do not overuse and underuse the definite article in a clear and definitive way. The category in which there is a suggestion that this may have occurred is persuasive writing (with press editorials serving as data). Collectively Outer Circle varieties' editorials may have used more definite articles than they would by chance. However, there are two arguments that can undermine this possibility in the bigger picture. First, the overuse may have been very minimal ($1.979567 - 1.96 = 0.019567$). Second, the Outer Circle group's frequency itself fell out of the normal range, tending towards underuse. While it, too, is very minimal ($-1.986316 - 1.96 = -0.026316$), the negative direction of this distribution prompts one to ask if the insertion and omission of the definite article are attributable to factors not yet considered here.

5.6 Results: *a/an*

The frequencies for the indefinite article per thousand words are shown in table 5.6. For the indefinite article, three types of comparison were made. The chi-squared test results for comparison (1) (i.e. the distribution of the frequencies in each variety) were not

significant.³² Comparing the frequencies from the perspective of text types (comparison (2)), the chi-squared tests also found that the differences between varieties in each text type were not significant except for private dialogue.³³ Using an alpha of 0.05, a post hoc analysis found that the variety which contributed to the significant chi-square value is AmE with a residual of 5.36719835. In the third comparison, the differences in the frequencies between the Inner Circle and Outer Circle groups were not found to be statistically significant ($\chi_2(9) = 2.5699, p > 0.05$).

³² The chi-squared values are as follows: Singapore ($\chi_2(11) = 7.2958, p > 0.05$), India ($\chi_2(11) = 16.8657, p > 0.05$), Philippines ($\chi_2(11) = 6.5941, p > 0.05$), Kenya ($\chi_2(10) = 4.0167, p > 0.05$), Britain ($\chi_2(11) = 5.4842, p > 0.05$), Australia ($\chi_2(11) = 4.7891, p > 0.05$), New Zealand ($\chi_2(11) = 17.125, p > 0.05$), and United States ($\chi_2(6) = 1.4653, p > 0.05$).

³³ The chi-squared values are as follows: public dialogue ($\chi_2(7) = 35.1939, p < 0.05$), public dialogue ($\chi_2(6) = 1.1373, p > 0.05$), unscripted monologue ($\chi_2(5) = 2.0204, p > 0.05$), scripted monologue ($\chi_2(6) = 5.0333, p > 0.05$), nonprofessional writing ($\chi_2(6) = 5.8012, p > 0.05$), correspondence ($\chi_2(6) = 2.855, p > 0.05$), academic writing ($\chi_2(7) = 2.3871, p > 0.05$), nonacademic writing ($\chi_2(7) = 1.7117, p > 0.05$), reportage ($\chi_2(7) = 2.8667, p > 0.05$), instructional writing ($\chi_2(7) = 3.9014, p > 0.05$), persuasive writing ($\chi_2(7) = 1.4718, p > 0.05$) and creative writing ($\chi_2(7) = 5.4356, p > 0.05$).

Table 5.6 Frequencies of *a/an* across varieties and text types

	S1A	S1B	S2A	S2B	W1A	W1B	W2A	W2B	W2C	W2D	W2E	W2F	Average
SIN	15	24	23	26	26	19	25	28	24	20	29	25	24
IND	16	20	25	26	14	13	19	29	23	28	25	30	22
PHI	18	20	20	21	25	19	24	26	19	26	23	30	23
KEN	20	22	NA	20	29	17	20	23	22	22	22	23	22
GB	17	20	23	24	24	20	26	29	27	27	23	25	24
AUS	21	25	28	30	27	22	22	30	24	31	26	27	26
NZ	13	22	28	33	26	21	27	31	28	30	23	38	27
US	45	NA	NA	NA	NA	NA	23	26	28	29	24	27	29
Average	21	22	25	26	24	19	23	28	24	27	24	28	

5.7 Discussion: *a/an*

With the exception of AmE it was found that the distribution of the indefinite article *a/an* is within the expected range for each variety. The discussion of AmE will be resumed below. As for the other varieties, on an individual basis, they do not demonstrate varying frequencies that allow us to speculate about their overuse or underuse of the indefinite article. This possibility is strengthened by the results of the tests conducted for comparison (2). Similarly the result of comparison (3) where the totals are amalgamated to produce a broad profile of the Inner Circle group and the Outer Circle group (excluding AmE and KenE) shows that there are no significant differences between them. It can be said then that neither group has overused or underused the indefinite article when compared to each other. This also means that there is no correlation that can be observed between absence of indefinite articles in L1 and overuse and underuse of *a/an*. The same is also true for a correlation between the influence of text types and overuse and underuse of *a/an*.

Having said this, AmE is clearly different in its significant overuse of *a/an* in the private dialogue category. It is important to remember that this result only emerges when the frequencies obtained in the other AmE text types are brought into comparison. This is reminiscent of the finding for *the* in 5.4. On scrutiny, there are only a few occurrences of repairs involving the indefinite article, so the same conclusion drawn for the definite article does not hold here. Since this is the second time an overuse has occurred in the same category, it is possible that the use of SBC as data representing private dialogue in AmE has led to this persistently discrepant finding.

There are perhaps some sampling criteria differences that exist between this corpus and the spoken component of ICE. However until ICE-US is completed and made available, this speculation cannot be confirmed. On the other hand the unexpected finding does point to the need for a qualitative analysis to complement the picture that has been obtained here.

5.8 Conclusion

In nativisation terms, this study has shown that omission and insertion of articles do not occur simply because there are inherited grammatical moulds from L1 waiting to be filled in, whenever available, with matching elements from English. Instead the use of articles is shown to be more complex than this.

In regard to the definite article, less proficient speakers in the Outer Circle may contribute to the characterisation of article use but this is limited to spoken language, and to a much lesser extent, correspondence. Overall, varieties are found to use the definite article variably and this was shown in their individual frequencies. This variability however underscores the fact that certain text types such as private dialogue have a tendency to use fewer definite articles than others. It has been suggested that this is due to the informality of the language contained therein. Only the private dialogue in AmE is shown to depart from this trend. It should be remembered that data from SBC are not thoroughly similar to those from ICE. Telephone conversations, for example, are missing from SBC. This, along with other characteristics discussed above, may have contributed to the frequency obtained for AmE. Furthermore, we recognise that a

comparison of this kind cannot identify which varieties exactly influence the variability. Thus, using text types as the basis of comparison, we saw that overuse is more common in NZE and underuse in AmE. Both are Inner Circle varieties. Nevertheless, further investigation found that the reasons for these phenomena are mostly text-internal and have nothing to do with the nativisation processes discussed in chapter 3. Consequently we submit that underuse does not necessarily equal omission and overuse does not necessarily equal insertion.

For the indefinite article, much less variation was discovered. AmE was found to deviate from the majority again in regard to its overuse of *a/an*. We postulate that this could be due to sampling differences between SBC and ICE. In other aspects, there are insignificant variation patterns found across the three types of comparisons. We conclude that, at least in terms of distribution, the use of the indefinite article is more stable across the varieties than the use of the definite article.

In conclusion, although there are slight tendencies to overuse and underuse the definite article in all the varieties, it is hard to attribute them to nativisation. If anything, we know now that the kinds of deviant uses in the Outer Circle varieties as illustrated in the literature (see chapter 3) are hard to find in naturally occurring data. This frequency analysis has effectively confirmed the subtlety of article variation in quantitative terms. To uncover it, the analysis must be increasingly qualitative.

Chapter 6 The Collocations of Articles

Collocation is said to be pervasive in language (e.g. Wray, 2002; Sinclair, 1991). This fact requires us to see that some units of meaning extend beyond the usual one-to-one correspondence between a word, either grammatical or lexical, and meaning. Furthermore it was argued that the boundary between lexis and grammar can be blurry. Building on the frequency study described in the previous chapter, chapter six describes articles from a collocational perspective. Section 6.1 describes an investigation of the collocates of *the* and *a/an* and examines some previous findings by Sand (2004) for *the*; 6.2 presents the findings for *the* while 6.3 presents the findings for *a/an* and 6.4 concludes the chapter.

6.1 A study of the collocational patterns of articles

Given that a word collocates with certain other words, the interesting question is: what can collocational patterns of articles tell us about their variation across the varieties?. The software used for this analysis is Wordsmith Tool 4.0 (Scott, 2004), which allows the computation of collocates of a search word in both left and right directions based on their relative frequencies. While all the corpora were investigated, the American variety is represented by only six categories from Frown that correspond to ICE's academic writing (W2A), non-academic writing (W2B), reportage (W2C), instructional writing (W2D), persuasive writing (W2E) and creative writing (W2F). As explained in chapter four, data from SBC and Frown enable some comparison to be made with the American variety. Because a more balanced comparison with this variety is not possible without

its own ICE component, any comparison that can be made with the available American data at this stage is useful as long as this is taken as an approximation only. Given this limitation, for wider comparability and applicability data from COCA and BYU-BNC were also used when information on native usage was lacking (see 4.2.3).

6.1.1 Collocates of *the*: some general observations

The first two-word analysis targets collocates of *the* in the first position to its right (henceforth R-1). However before we commence discussion of the differences, it is worthwhile to make some general comments on the similarities observed across all corpora.

Some nouns and adjectives are consistently highly ranked in the R-1 position i.e. appearing in almost equal proportions in all varieties. Figure 6.1 shows the nouns that are found in the top 20, 50 and 100 words in all varieties.

Figure 6.1: Most frequent noun collocates of *the* at R-1 position in all varieties
(except US where noted)

Top 20	Top 50	Top 100
1. way	1. government	1. people
2. time	2. public	2. house
3. world	3. problem	3. past
		4. country

There are more nouns shared by the varieties beyond the top 100 words but this analysis focuses on the ones that appear in the top 100 only because they are the most salient.

Why do these nouns collocate with the definite article the most? At first glance, no clear pattern emerges from this small group of nouns. Looking at the majority of them, however, they confirm the findings of Fraurud (1990) and Poesio and Vieira (1998) that, despite anaphoric reference being cited as the most common usage of the definite article, its most common determinant is in fact reliance on situational context (see table 2.1). The NP *the world* in top 20, for example, requires general knowledge in a larger situation (92% of all the tokens). This is also true for some other words in figure 6.1 i.e. *government* (78%), *country* (63%) and *problem*, although it was found that *problem* can also be indirectly anaphoric (55%). *Public* and *people* are substantially used to refer to ‘the masses’, this referent being mostly inferrable from the immediate and/or larger situation (98% and 51% respectively).

House can be inferred from the larger situation when it refers to the parliamentary House of Commons or House of Lords in the British Commonwealth varieties or House of Representatives in the Philippine variety (25%) as well as when it refers to a known house in a larger situation (39%) as in this example:

71. "What the hell is happening Matthew?" Mother demanded, the pearls around her cheongsam collar gleaming softly; seeming to reflect the heat of her

agitation, " We go out for one evening and the house is a war zone... <ICE-SIN:W2F-007#162>.

It may also be used anaphorically with *the* (36%) as in:

72. I should declare a personal interest in the proposed development in that The Old Rectory will be one of the houses most directly affected. The house is not a listed property (we have always felt grateful for this) but selection for listing in this area has been curiously arbitrary, based exclusively on whether a building is in stone <ICE-GB:W1B-020 #119 - 120>.

The other top 20 words *way* and *time* are more complicated. Consulting the concordances of these words in all the corpora, one finds that when they occur with *the*, they are not commonly used in simple NPs such as *Can you show me the way?* or *Can you tell me the time?* Rather, they are metaphorically construed and occur in various expressions with an idiomatic tendency (68%), e.g. *by the way*, *all the way*, *in the way of*, *along the way*, *on the way*, *all the time*, *the time being*, *half the time*, *most of the time*, etc. Both words are also frequently postmodified (53%). Examples are given below:

73. That that's college I mean sometimes you just have to say that that that's the way that's the <.> wa </.> that's the way things happen here <ICE-PHI:S1A-075#59:>.

74. If you are at *the time of life* when your bone mass is declining, exercise can slow this process down which calcium on its own can't <ICE-GB:W2B-022 #43>.

It could be argued that some special properties of these words, i.e. ability to transform into metaphors, idiomaticity and propensity for postmodification, may contribute to their high frequencies.

The remaining word, *past*, may at first appear to refer only generically to a time dimension but its concordances indicate that nearly half of them (42%) are used as adjectives as in:

75. The Seven News Morgan Exit Poll shows an Election eve swing to the Coalition, the ALP losing two vital points in *the past twenty four hours* <ICE-AUS: S2B-007(B):72>

leaving the NP of interest *the past* at about 58%.

Other than the fact that the majority of nouns used with *the* in all varieties are not anaphoric, it can be said that there are no easily distinguishable factors that explain their frequent collocations. Perhaps the best explanation is that the concepts that these nouns denote are frequently important to human experience. It must be pointed out that the majority of the nouns in figure 6.1 have to do with large-scale human institutions (or in the case of *world*, an entity that concerns all human beings). These include *government*, *public*, *people*, *house* (in the parliamentary/congressional sense) and

country. This connection could partially explain the words on the list in the above figure. A rather simplistic piece of evidence for this claim can be offered through an analysis of less frequent collocates of *the* in the varieties. Examining R-1 collocates of the definite article with a frequency of ten, we find that there are no commonly shared nouns at this level of delicacy in *all* varieties. The nouns that are found here are less readily applicable to such large-scale institutions. Some examples of these include *debate* (ICE-GB), *welfare* (ICE-NZ), *seat* (ICE-AUS), *survey* (ICE-PHIL), *leg* (ICE-IND), *drug* (ICE-KEN) and *bed* (ICE-SIN). Now we shall examine the other major part of speech found in the R-1 position, i.e. adjective.

Unlike nouns, there is a large number of overlapping adjectives between the varieties. Figure 6.3 shows the adjectives that are found in the top 20, 50 and 100 words in all varieties. Words that can function equally as noun or adjective were excluded. Some examples are *top*, *left*, *end*, *one*, *bottom*, etc.

Figure 6.2: Most frequent adjectival collocates of *the* at R-1 position in all varieties

Top 20	Top 50	Top 100
1. same	1. next	1. main
2. first	2. only	2. following
3. other	3. right	3. third
4. last	4. new	4. united
	5. best	
	6. second	

In contrast to the nouns, it is easier to understand why these adjectives are top-ranking collocates of *the*. Except for *united*, all the other adjectives are semantically compatible with definiteness, making them suitable candidates for *the*. This is not to say that these adjectives cannot occur with *a/an* or \emptyset (except for *same*; cf. * ‘a same shirt’) but they lend themselves very easily to definiteness as compared to adjectives which are found at the lower end of the ranking, e.g. *Asian* (ICE-Sin), *armed* (ICE-Ken), *proposed* (ICE-Ind), *alleged* (ICE-Phil), *open* (ICE-Aus), *hard* (ICE-NZ), *larger* (ICE-GB), etc. Incidentally, *united* is frequently part of the phrases *the United States*, and to a lesser extent, *the United Kingdom*, both indicating the global significance of the two countries in the English-speaking world.

These are the similarities observed across the corpora. Differences that apply universally to either the Inner Circle group or the Outer Circle group are harder to find.

There are two cases that approximately fulfill this condition. The first is the word *top*, which is found in all Inner Circle varieties but in only two Outer Circle varieties. It is found particularly in the top 40 words of the Inner Circle varieties but at a lower rank in the two Outer Circle varieties (at position 51 in ICE-Sin and 154 in ICE-Ind). *Top* is used as an adjective, as in:

76. *The only top placings* since then have been third through seventh at Tampa in 1989. In 1990 and '91 no one made *the top six* <E09 72 – E09 74; FROWN>;

or as a noun, as in:

77. It was quite incredible the way we got the case of beers over the fence. It was all locked off. Yeah . Climbed over *the top* and got the the beers through and all the rest of it < ICE-AUS: S1A-031(B):175 – 178>.

Less commonly, it is used in idioms such as:

78. In his most recent recording for DG, Bernstein rather goes *over the top* with his jazzing of the solos in Gershwin <ICE-GB:W2B-008 #88>.

Apparently *top* is not as widely used in the Outer Circle varieties, especially in PhilE and KenE. It is possible to speculate on the causes for its lack of use as a noun and in idioms. Using (77) as an example, it is possible that speakers of Outer Circle varieties prefer other constructions to *climbed over the top*, such as *climbed over it* (the fence). The motivation for this preference is difficult to ascertain. However, as suggested by some linguists (e.g. Schneider, 2007; Ansaldo, 2004; Gil, 1994), linguistic items not available in L1 may be avoided by speakers and this is possibly true for the 'top of the fence' which is usually regarded as a constituent meaning of 'fence' in at least Chinese,

Malay, Tagalog and Tamil. *The top* is already understood as present when *fence* is uttered. An example of a substitute for an idiom is when a speaker of an Outer Circle variety chooses *Bernstein overdoes his jazzing of the solos in Gershwin* instead of the original in (78) (although the effect may be slightly different). If these speculations are valid, the frequency of use of an article is then also tied to whether or not a lexical item is preferred by speakers. Notice that if a speaker uses *it* instead of *the fence*, then *the* is not used. At present, it is apparent that the use of a particular word such as *top* by speakers of Inner Circle varieties is, in the final analysis, a minor variation of the definite article because it is merely a case of preference.

The other case in which a word is found in only one group of varieties is more symptomatic of nativisation, in particular the word *society* (in the non-club sense). This word is found among the top 500 words of three out of four Outer Circle varieties (at position 39 in ICE-Ken, 106 in ICE-Ind and 414 in ICE-Sin). This means that the NP *the society* is frequently and substantially used by speakers of Outer Circle varieties (excluding Phile). It is a pattern of use that was also highlighted by Sand (2004: 292) and Mair for Jamaican English (2002: 51-52). But what makes *society* collocate with *the* so commonly in the Outer Circle varieties? One reason could be that the meaning of *society* is at times interchangeable with ‘community’, which in fact requires the definite article. Viewed as such, it is understandable why *society* needs to be determined by *the*. It, too, is a (loosely defined) collective group of people but one that needs to be marked for identifiability. (An alternative interpretation, based on the data of this study, is

given in 8.2.2.) Nevertheless, as we shall see, there are very few nouns like *society* which receive definiteness marking solely by virtue of their internal meaning.

In her study, Sand also argues that that the definite article is used in non-native varieties to (1) signal generic meaning in certain collective nouns, such as *society* above, (2) accompany “nouns referring to institutions”, (3) determine “the proper name of a place or institution [when] British and American standard English do not use the article” and (4) occur with temporal expressions without the accompanying postmodification (2004:290-291). These non-nativelike uses were identified through their violation of some of the rules of articles laid down in Quirk et al. (1985). However, based on the collocates of *the* found in the data, it seems unwise to tie the unconventional marking of definiteness for these NPs to the lexical semantics of the nouns alone. In what follows, I will try to explain in considerable detail to why the majority of these supposedly untypical uses are not acceptable as variation patterns particular to Outer Circle varieties. The primary motivation for this effort is the belief that the use of an article is tied to the interaction of its semantics *with* that of the NP’s head. It will be shown that, while certain nouns identified by Sand are sometimes unconventionally marked for definiteness with *the* by Outer Circle varieties’ speakers, the nouns themselves do not exclusively contribute to that definite meaning. This implies that those nouns can felicitously occur with *the* – as they do in Sand’s examples – under different circumstances.

Sand's examples for the use of *the* with collective nouns are listed in (9) below:

- 79 a. *The girls* tend to fair (sic) better in these subjects <ICE-SIN: W1A-007>.
- b. But that is one problem *the people* don't believe in homeopathy still <ICE-IND: S1A-037>.
- c. Like, *the people* weren't so particular that time <NITCS: Belf111>.
- e. And *the people* have thought that philosophy must be uh it creates trouble <ICE-KEN: broadcast discussion>.
- f. ... I couldn't help wondering once again what it is that gives *the girls* such an advantage over *the boys* <ICE-JAM: W2E-008>.

If we look carefully at (79) again, the use of *the* is not always entirely odd. These words (*people*, *boys* and *girls*) are found to be regular R-1 collocates of the definite article in *both* Inner and Outer Circle varieties. In particular they are found among the top 500 R-1 collocates of *the*. Although reference grammars may present collective nouns as typically determined with a zero article for genericity (e.g., Quirk et al., 1985; 288) there are a number of instances in which *the people* appears to be used in a generic sense. These examples were taken from BrE and AusE:

80. We want ongoing public participation so that fundamental changes to communities cannot take place if *the people* don't want them <ICE-GB: W2B-013 033>.
81. The union members are effectively snubbing their noses at *the people* as a whole with the results of their actions reflected in higher prices, supermarket

shelf shortages and increased costs to exporters <ICE-AUS:W2E-010 (noone):25>.

What does this suggest? On one level, it seems that even native speakers sometimes use a definite article with a collective noun to mark genericity. Perhaps what Sand has demonstrated is that such NP's are more frequent in Outer Circle varieties. On another level, we might surmise that the pragmatic context may sometimes warrant a definite marking for a *seemingly* generic NP, although this may not been evident without a larger context. Indeed this appears to be the case when one looks at (79a) – (79b). Consider (79a) again. When seen in its larger context, which is provided here, there is evidence to believe that the referent of *girls* is a special subset singled out for identifiability:

The meritocratic system which stresses on academic qualifications, and result and achievement oriented attitudes, which leads to few opportunities for young men and women to meet since they deem it as more important to concentrate on getting good results in their academic examinations and work excellence. The problem is especially acute during schooling days, which is the best period for them to forge a lasting relationship. Also, an average 63% of boys who did well enough at 'O' levels chose to enter polytechnics instead of college/pre-university classes. These boys are apprehensive of the second language and general paper requirements for entry into the National University of Singapore, the only university at the moment. *The girls* tend to fair (sic) better in these subjects <ICE-SIN:W1A-007#40-44>.

It is true that the use of *girls* (without the definite article) would be more expected here. However it is not difficult to appreciate an interpretation where the girls the writer refers to are the identifiable, and thus definite, female counterparts of the males who did well enough at ‘O’ levels.³⁴ These girls too did well enough in this exam but were not apprehensive about the second language and general paper requirements for entry into the university.

Sand’s examples showing the definite article being used in cases “when a type of institution rather than a specific ... body is referred to” (2004: 290) are given in (82) below:

- 82 a. We don’t need to go to the university to acquire you know a degree in mathematics simple arithmetic will get you the answer <ICE-SIN S2A-022>.
- b. (How did you learn Hindi?) So Hindi was compulsory for us uh in the school <ICE- IND S1A-076>.
- c. And her eh sister’s going over to Scotland to the university <NITCS Belfl 19>.
- d. So you must be wondering why people must go to the church <ICE-EA(K) conversations>.
- e. Also Jamaicans would not have to enter the university in the prime years of their adult lives...<ICE-JAM W1A-015>.

While the above examples seem to indicate that such is the variation in the use of *the* in Outer Circle varieties, as with (79) it would be worthwhile to consult (i) each NP’s

³⁴ That they are anaphoric to the antecedent *young women* is another possible way to explain *the girls*.

larger context and (ii) the patterns of similar NPs in Outer Circle varieties to confirm their cases.

The use of *university* in (82a) would seem more typical if it occurred with zero but there are cases of *the* being used with a type of institution in the Inner Circle varieties.

Such use is attested several times at least in the NZE data:

83. Perh</.> perhaps the course <.>sh</.> doesn't belong in the university perhaps belongs in a um a polytech or a um you know one of these [,] things you do in three days<ICE-NZ:S1B-007#286>.
84. I think baldwin's learning society has been one of the ideas that's influenced me most in the last year or so and ball's idea I see the university as a cardinal component of that learning society <ICE-NZ:S2A-038#161>.
85. I think *the university* is then a fundamental component of the <.>unac</.> human capital creation of any society <ICE-NZ:S2A-038#129>.

If Outer Circle varieties do make use of *the university* generically, then it is not different from their Inner Circle counterparts in absolute terms. This could also explain (82c) but we do not have the Irish data to confirm its similarity to (82a).³⁵ Having said this, it needs to be pointed out that the phrase *go to the university* with its seeming generic reference as found in (82a) is not a typical native use. This phrase and its meaning was checked with the 300-million word COCA and the 100-million word BYU-BNC since the amount of data in the native varieties of ICE turned out to be

³⁵ It is also possible that 'the university' in (82c) is a specific referent.

lacking in this sense. COCA yielded 84 phrases and BYU-BNC produced three. Out of these, three appear to be generic, or at least as generic as Sand's examples appear to be. All three are from COCA.

86. ...over the beautiful silver skates and flowers on the table. Gretel and Hans, arms entwined, smiled as their father rose from his chair, snapped his fingers, and turned on his heels as if to dance. # Although Hans had not won a prize, he had captured the respect of Dr. Boekman. On the great doctor's next visit, he asked the boy an important question. "Would you like to become a physician, Hans?" # " Yes, sir," Hans said eagerly. # "Would you go to the university, study very hard, and become my assistant?" # "I would, sir." # "You would not change your mind as I prepared you to be my successor?" # "No, sir." <COCA:2006/MAG/SatEvenPost>.

87. ...care of the elderly in the United States, we clean the rooms in the hospitals. But then when we have health needs, we're not able to get those met. And so we talked about the health needs of the Mexicans abroad. We also spoke about education. I commented to him about Rick Perry, governor -- also a Republican governor of the state of Texas; did sign a bill which is historic, the first state in the nation now that permits all those who graduate from high school to now be able to go to the university whether they can prove legal residency or not. Some of these are some of the brightest kids in the United States, but then could not go to junior colleges and colleges <COCA:2001/SPOK/NPR_ATCW>.

88. ...was I think more vague and there was so much more anti-war sentiment then than there is now, and I think it's a lot different. This may change if it's dragged out a long time, but right now, it looks like the American people are behind it, the military is eager to help if they're needed, and I think it's just a different attitude on the part of all of us than it was during Vietnam. HILLARY CLINTON (D-NY), PRESIDENTIAL CANDIDATE It feels good for me now to go to school now, because I go to the university to be in my uniform. I don't feel like a students (sic) are jeering me or anything like that. I can be a lot more proud of what I'm doing <COCA:1990/SPOK/PBS_Newshour>.

With this evidence, it can be said that (82a) is not unique to SgE, or Outer Circle varieties only. *The university* in (82e) seems to be slightly different. Here, the phrase is headed by the verb *enter*, not *go (to)*. It can be argued that the verb *go* has a more concrete sense than *enter*. Is it possible that the result is different for this verb? In COCA, it was found that *enter the university* with generic reference was used five times in the 18 phrases yielded. In BYU-BNC, however, none of the four phrases had this meaning. An example from COCA is given below:

89. Given the present climate of academia and the emphasis on scholarship, academia may be forcing a particular identity upon junior faculty, an identity that may be at odds with who they truly are. Not everyone enters academia from a desire to write. Some do enter the university from a desire to teach or for the intellectual stimulation of both. So, how best for junior faculty members to survive in a culture that may be shaping their identities in ways opposed to their own self-perceptions? <COCA:2001/ACAD/Education>.

If the pattern in (82e) and its intended meaning are also not unique to Jamaican English, and hence, Outer Circle varieties, then we are left with (82b) and (82d) as evidence for Sand's claim. In (82b), however, notice from its larger context – provided here – that the NP *the school* was not a direct continuation from where the filler *uh* was uttered by speaker A.

82bi. B: No how did <,> how did you learn Hindi <,>

A: So Hindi it was compulsory for us <,> in uh

B: In *the school*

A: Yes yes in *the school*

<ICE-IND:S1A-076#111 – 114>

In fact, this NP was first supplied by speaker B and subsequently taken up by A. In the extract's even larger context, B was found to be interviewing A with the latter informing his interviewer of his place of origin (i.e. the Bombay-Karnataka region). As the topic of the conversation was A's ability to speak Hindi despite his Marathi-speaking background, it is possible that the mention of this place triggered region-specific institutions, including schools, for the speakers. In other words, *the school* is likely to be a specific referent for the speakers (first by B, then by A) despite its generic appearance. In chapter 7, evidence for the nativised use of *the* as a marker of specificity by the speaker will also be provided.

What can be said about *the church* in (72d) then? This is the closest example to the generic reference as noted by Sand. It seems from the closer inspection of its larger

context that speaker C's uses of the generic *the church* and \emptyset *church* are in free variation.

82di. C: Yeah so I was talking about this missionary in my place that he was telling me that he comes from Canada I just want to visit him in his place but he lives in Nairobi

B: Uh

C: so he told me <_>he told me</> that you see in Canada people don't care about going to the church^a

B: Uh

C: because people they have basically everything for human existence that they need like a car almost everybody is having a car

B: aha

C: and there is just no much problems

B: Uh

C: and some people will find no reason of going to church^b

B: Uh

C: So I told him so [,] people are materialistic in your place

B: Uh

C: He said Yeah kind of kind of people are materialistic but you see

B: Uh

C: Yeah somebody can have that concept of God and you can know that God is there

B: Uh

C: but to bother to go to church^c or why go to church^d there is no need

B: Uh

C: unless

B: There are no problems

C: Yeah unless one is in full service that is the time he is going to take himself now to the church^e for

B: Uh

C: such purposes

B: Uh

C: So I was trying to compare it with Africa so he said now in our case if you come to Africa so how do you find here

B: Uh

C: So he declined to comment

B: aha

C: Sometimes I don't know due to personal reasons or what he declined to comment the

B: He is a foreigner isn't it

C: Yeah he is a foreigner

B: Uh

C: But he was very good he was very social

B: It could be [,] it could be [,] he was declining because he does not know much about Africa

C: Yeah maybe because he was enquiring a lot of things from me

B: Uh

C: So

B: And for that matter what <-/>what was he implying Was he implying
that in Canada uh people are not

B: people are not all that religious

C: Yeah

B: And then if they are it is just an insignificant of them

C: Yeah

B: Okay

C: Yeah they don't see it because people are comfortable

B: I see

C: So you bring a [...] to them tell them that this is the new way

B: Uh

C: And they are already comfortable you see uh

B: They don't bother

C: So the impact is just insignificant

C: So to a very big extent or to a large extent

B: Uh

C: What <O/> was saying that uh

B: Uh

C: You see when you have problem is the time people rush to church^f I
think is a

C: is appointment to

- B: Uh
- C: to hold water
- B: Uh
- C: Uh
- B: So you must be wondering why people must go to *the church*^s
- C: Yes that is the question I have been asking myself for a very long time
- B: mhm

<ICE-KEN: S1A016K>

It is not known if speaker B also uses *the church* and \emptyset *church* interchangeably for generic meaning. Sand's example of generic 'the church' is (g), which occurs only once. As with other examples in (72), Sand's evidence for the so-called violation of article usage rules in nativised varieties is essentially impressionistic and not quantified. As regards generic *the church* in ICE-Ken, it was found that there was only one such case in the whole corpus (again the search is narrowed down to the phrase *go/going/gone/went to (the) church* for the same reason that this phrase is untypical for generic reference in Inner Circle varieties). Similar use of *the church* cannot be attested in the corpora of the other nativised varieties. We can then conclude that, while some speakers of KenE may use *the church* for generic meaning, it is not widespread in that variety and is not shared by speakers of other Outer Circle varieties.

The use of *the* to accompany "the proper name of a place or institution" is illustrated by Sand by the following:

- 90a. Thankfully, he was sent to the Singapore General Hospital early enough <ICE-SIN:W1B-009>.
- b. ...- says Amarty Sen, Lamont University Professor at the Harvard University <ICE- IND:W2B-015>.
- c. And there was another place, I think, was in the county Down, somebody had them, and they ploughed with them <NITCS Belf13>.
- d. ...the Panadol Hockey Tournament which started yesterday at the Jamhuri Highschool continued today at the same venue <ICE-EA(K) news broadcast>.
- e. In 1987, the Victoria Park was transformed into a thing of beauty <ICE-JAM:W2C-017>.

At first glance, (90b) and (90e) are the most untypical. This is because the names of universities, schools and popular parks are already unique denotations (cf. Quirk et al., 1985: 288). Thus we typically have 'Ø Oxford University' and 'Ø Hyde Park'. A straightforward rebuttal of this point would be to point to cases such as the following:

- 91. Plans for the unitech began in 1979 but negotiations started last year between Te Runanga O Ngati Awa - representing the Ngati Awa people - and the Waikato University <ICE-NZ:W2C-017#22>.
- 92. We've been thrilled with this work experience cos was Dave originally had these five days off taping and he said no, he said, you know you're so tired, have a little rest we'll go up I'll go up on the train only to get to the Selhurst Park he's got four or five train journeys Yeah mm and he has to get, like Woking to Surbiton [,] Surbiton to Clapham Junction <BNC-BYU:KBF/S_conv>.

However a more serious shortcoming of the claim relating to the use of *the* with proper names must be pointed out; namely (i) the existence of universities that take *the* as part of their names³⁶, and (ii) the fact that uniqueness is not universal. To illustrate (i), consider the name of a quite well-known medical teaching institution, ‘The Rockefeller University’ as in:

93. If you're just joining us, my guest is Dr. Bruce McEwen. He's the head of the neuroendocrinology lab at the Rockefeller University in New York City <COCA:2002/SPOK/NPR_FreshAir>.

94. Knowing the 3-D structure of a molecule is critical because “in biology, there is an intimate link between form and function,” says Dr. Stephen K. Burley, head of the Laboratories of Molecular Biophysics, and a Howard Hughes Medical Institute investigator at The Rockefeller University, New York <COCA: 1993/NEWS/USAToday>.

Occasionally the use of *the* is extended to universities whose names do not include the definite article but the result is more natural-sounding than *the Harvard University* in (90b). This can be seen in:

95. No. It really didn't. We've got the Ohio University here and because it's such a great place to live, the market has been great for real estate. Unfortunately, I came from an industrial city and in industrial-manufactured city, the housing prices are just heading for the cellar <COCA: 2008/SPOK/NPR_TalkNat>.

³⁶ These do not include those with ‘of + name-of-place’ PP, e.g. The University of New South Wales

To understand why the alternation between ‘Ohio University’ and ‘the Ohio University’ is successful, and that between ‘Harvard University’ and ‘the Harvard University’ is seemingly not, requires us to consider the second reason listed above, i.e. uniqueness is not universal.³⁷ To elucidate this argument, let us examine the use of ‘Yellowstone Park’ and ‘Yosemite Park’, two well known American national parks. None of the approximately 140 NPs denoting these entities occurs with *the* in COCA. This is unsurprising given the familiarity of the parks among many Americans. Consider now two more parks of more or less the same orientation. One is the lesser known Adirondack (a state park) in the state of New York and the other is the Yanomami Park (a proposed reserve for the Yanomami people) in Brazil. The Adirondack Park, it seems, can be found both with and without the definite article:

96a. The Adirondack Mountains are a curious secret. # Stretching across more than 6 million acres that dominate the upper half of New York state, Adirondack Park would seem hard to miss <COCA:1995/NEWS/Atlanta>.

b. On the other hand, Robert C. Glennon, the APA's executive director, sees disaster looming. “The scheme of the present zoning law guarantees the ultimate destruction of the Adirondack Park and anoints us to preside over it,” he says <COCA:1992/MAG/NatlParks>.

The alternation between *the* and zero in the sentences in (96) is successful too. Why is this so? It appears that the park in question is so well-known to some speakers that they are able to recognise its uniqueness at the mention of its name without *the*. This is the

³⁷ Another well-known public university in Ohio officially takes the definite article: ‘The Ohio State University’.

case with (96a) but not (96b). Let us provide further evidence for the uniqueness factor. The Yanomami Park, in contrast to the Adirondack, is arguably unknown for the majority of Americans and its uniqueness is thus not readily established. It can be found only once in COCA³⁸ and occurs with *the*:

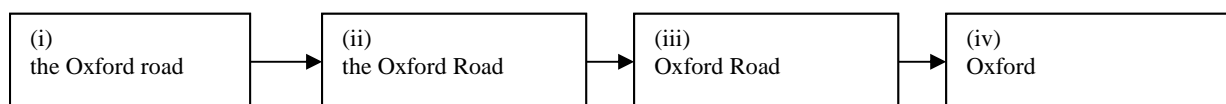
97. Claudia Andujar, coordinator of the not-for-profit Commission for the Creation of *the Yanomami Park* (CCPY) says the Yanomami issue is a "test case" that will show how Brazil intends to treat its indigenous peoples and their environments <COCA:1991/NEWS/CSMonitor>.

Having established the relative strength of uniqueness in the use or non-use of *the* to accompany proper names, let us return to the cases of (90b) and (90e). Although it may come as a surprise for some people, it is possible that for the speaker of (90b), the NP *Harvard University* is not sufficiently unique to warrant a proper name treatment. The same can be said of the speaker of *the Victoria Park* in (90e). Although it can be argued the fact that the park is located in Jamaica should provide some familiarity and turn the park into a unique entity for this Jamaican speaker, we saw in (96b) that this cannot always be guaranteed. In (96b) the speaker too shares the same country of origin with the park concerned but uses the definite article while the one in (96a) does not. Quirk et al. explain this alternation in terms of the “gradient” between proper names and definite NPs as illustrated in the following diagram:

³⁸ However in BNC-BYU, it occurs six times, all with ‘the’. Here is an example:

In a surprise move, Brazilian President Fernando Collor de Mellor has issued a decree establishing a reserve for the Yanomami people on their traditional lands in north-eastern Brazil. *The Yanomami Park* will cover 9,400 million hectares on the border with Venezuela <BNC-BYU:J32/W_Misc>.

Figure 6.3: The gradient between proper names and definite NPs (adapted from Quirk et al., 1985: 294)



According to them, (i) the NP is definite; (ii) the definite NP is “conventionalized to the extent that *Road* is capitalized, so that the whole makes a composite name”; (iii) article deletion is a sign of further conventionalisation, “[t]he purest example of a name is perhaps (iv)” (ibid.). We see now that cases like ‘Harvard University’ and ‘the Harvard University’ are still fluctuating between (ii) and (iii). Importantly this fluctuation is visible to observers because speakers from different cultural contexts such as those of Inner Circle and Outer Circle varieties are involved.

Now we turn our attention to the rest of the sentences in (90). After examining (90a), (90c) and (90d), we find that the evidence is still not sufficiently convincing to say that these are variation patterns that are unique to Outer Circle varieties only. As regards (90a), Quirk et al. point out that the names of public institutions such as hospitals are “typically” preceded by *the* (1985: 296) as in:

98. And on the twenty-seventh of March, eighty-seven, she was readmitted to the Queen Victoria Hospital, and the implant was removed, under general anaesthetic because it had become infected <ICE-GB:S2A-062 #46>.

Cases like ‘Ø Queen Victoria Hospital’ are present too in the British data and I suggest the uniqueness factor discussed above is at play here. The same can be said about (90d).

The following is an example of such Inner Circle usage from COCA:

99. Six months after graduating *The Woodlands High School* in 1998, Gonzalez enlisted with the U.S. Marine Corps <COCA:2007/NEWS/Houston>.

As for (90c), we can find an identical construction in BYU-BNC. The NP identified by Sand and the one in (100) below have an unmistakable Irish influence. The latter is in fact found in a text of Irish origin:

100. Patrick's mother had died twenty years before in *the County Donegal*, trying to prove that she could still climb to the top of a mountain called Muckish <BNC-BYU:HTJ/W_fict_prose>.

Nevertheless, since (100) is found in a corpus of British English, we can effectively disregard Sand’s claim for (90c). Given also Hickey’s claim that “[English] is undoubtedly the native language of the vast majority of the [Irish] population” (2007: 22), it is felt that, where the distinction between Inner and Outer Circle varieties is concerned, this decision is justified.

Sand also claims that an untypical use of *the* in Outer Circle varieties is found with “temporal expressions, like days, months, seasons and holidays” (2004: 291). Her examples are given below:

101a. *the Saturday* was my last one <ICE-SIN S1A-001>.

- b. but the trees burst into buds and put on fresh one when *the spring* comes around <ICE-IND S2A-025>.
- c. I'm going over the meeting on *the Thursday* <NICTS Belf17>.
- d. The most universalized celebration is *the Christmas* ... <ICE-EA(K) exam essay>.
- e. Leave them for now let them enjoy *the Christmas* <ICE-JAM S2B-016>.

In (101) we do see that (a), (b), (d) and (e) are all instances of more or less marked use of the definite article in the Outer Circle varieties. However inspection of the relevant concordances in the Outer Circle corpora revealed that there are less than five instances of such uses altogether. As regards (101c), Quirk et al. point out in their footnote that “[i]n rather popular BrE usage (felt by some to be non-standard), the days of the week have the definite article in “And on *the Thursday* she got worse. So on *the Friday* we called the doctor [‘the Thursday, etc of the week concerned’]” (1985: 293). The following is taken from ICE-GB:

102. So supposing when you come back on Sunday or say *the Monday* or something supposing you come in and get the thesis, would you feel able to have at least a look at it in that week and then give it back to me <ICE-GB:S1B-012 #215>.

It needs to be pointed out, however, that the use of ‘the + name of day’ is only felicitous when it collocates with *on*. Altogether there are five such cases out of 433 relevant NPs in ICE-GB.

Finally, Sand notes (p. 291) the Outer Circle varieties' tendency to use *the* with “deverbal nouns in –ing ... instead of verbal expressions.” Her examples are:

- 103a. I think it's moving that way beyond *the caring* for the general welfare and uh rice bowl <ICE-SIN:S1B-040>.
- b. We present a new concept in *the broadcasting* in India <ICE-IND:S2A-007>.
- c. When they're wet, they're stuck together, you know, but *the drying* separates them <NICTSBelf122>.
- d. So *the parking* is becoming a problem because of you know there are so many buildings there are so many vehicles <ICE-KEN: broadcast interview>.
- e. On came *the groaning* and *the stomping*, like a hundred men stepping on his chest to cut off his breath ...<ICE_JAM:W2F-003>.

In (103) we see what may appear to be more problematic usages of *the* according to native standards. Native speakers would only rarely say something along the lines of (103a), (103b) and (103d). This will be proven below. Nevertheless, it must be said that (103c) and (103e) do not appear to be as problematic as previously thought since, in fact, there are similar usages found in ICE-Aus and ICE-GB:

104. It's up to those who have a consciousness to talk about it and I think you have to encourage people to go because it's only in *the going* it's only in the confrontation it's only when the two world's meet and and clash that you get an interest and a discussion < ICE-AUS:S1B-047(B):206>.
105. Now everyone says oh well then it's an editor's film. Well of course *the editing* is extremely important <ICE-GB:S1B-045 #126-127>.

If the argument is that these deverbal nouns are more typically realised as verbal expressions, then we see that it is only applicable to (103c) but not (103e) (cf. ?*On came groaning and stomping....*).³⁹ The same argument may well be applied to the Inner Circle examples of (104) and (105) and we see that *the editing* in (105) can be changed to ‘editing’ but *the going* in (104) cannot be changed to ‘going’.

Coming back to (103a) and (103b), here we actually see a NP with a PP as postmodifier. In the Inner Circle varieties, and – importantly – in large parts of the Outer Circle data, the preposition commonly used in such PPs is *of*. Far less commonly other prepositions such as *to* are used. The former is exemplified in the first sentence of the example below; the latter in the second sentence.

106. First came the torture, *the scourging* and *mocking the forced carrying of* one's own heavy cross or maybe a cross-piece for which Jesus was evidently too weak to complete the course. Then *the nailing to the cross* <ICE-GB:S2B-028 #46 - 47>.

In the light of (106), the example in (103a) does not appear too convincing. (103b), however, remains semantically non-nativelike despite the grammatically formed NP. Nevertheless, this is the only one of its kind in all of ICE-IND and we will not take it as sufficient evidence of Sand’s claim.

³⁹ But the term ‘deverbal noun’ as used by Sand is confusing. If (103c) is realised as “...*but drying separates them*”, ‘drying’ by itself is a gerund but still very much a deverbal noun.

(103d) illustrates most clearly the problem of attributing the unconventional use of *the* to a single class of nouns that is based on their deverbalisation. It is a good example because it shows the speaker's otherwise hidden motivation for the use of highly untypical *the parking*. When seen in its larger context, notice that *the parking* arises from a context that has been set up prior to it within the same stretch of discourse. After talking about the crowding of the city where the traffic gets worse, speaker B uses *the parking*. It is rather untypical but it does give the meaning it is intended to convey.

A: People working in the city centre at times they have a problem of parking and they'll park anywhere maybe Corporal A you can tell us where to park because some people really have a problem in parking

B: Yes the problem is one Everyone wants to come to the city centre uh forgetting that we are now cumulating ourselves <O/> Okay the human traffic the motorists when they come together you'll find the condition of that movement is becoming thick So the parking is becoming a problem because of you know there are so many buildings there are so many vehicles there are so many human traffic So this a problem Otherwise maybe that uh we shall be sending our city to Machakos areas like that so that we'll get enough parking At the moment well the problem is not for a person alone it is for a problem to better take care We better obey the laws not causing obstruction by parking </anyhowly> you better get to uh bay parking bays or places set aside for a vehicle to be parked Otherwise the problem is very <?/>wide

for everybody not only for police officers or...

<ICE-KEN:S1BINT1K>

In other words, *the* is used here as in the context of the referent's larger situation (see table 2.1). And, as further evidence to the untenability of the claim of deverbal nouns providing the basis of untypical use of *the*, notice that the same speaker makes use of \emptyset *parking* three sentences later.

As a conclusion, variation patterns that are identified by Sand, although they have been discovered in the Outer Circle varieties, are not infallible. It would be hasty to conclude that variation in the use of *the* is exclusively defined by the properties of the nouns themselves. First of all, in addition to the fact that they are not entirely absent in the Inner Circle varieties, we saw that the frequencies of these putative variation patterns are small in the Outer Circle varieties. Further, the factors leading to the use of the definite article, though not always apparent at first, can be identified upon scrutiny of its context of use. What may be a more satisfactory account is the use to which the definite article is put. We saw such uses for the majority of the examples given by Sand. We are now also in a position to appreciate that the semantics of articles itself, as opposed to the meanings of nouns alone, has much to contribute to whether or not an NP takes a certain article. Nevertheless it is important to note that the pattern *the society* is found to be exclusive to the Outer Circle varieties. Far less frequently we find usages of *the Christmas*, *the spring* and the generic *the church*. These few NPs can then be said to be genuine patterns of variation in the Outer Circle varieties, with the caveat

that the latter three are not as widespread as *the society*. As described by Sand: “[t]he situation is less conclusive for other lexemes under analysis” (2004: 292).

We may hypothesise that the uses of articles boil down to their usage types. As mentioned above, these appear to be valid predictors for their variation. Recall *the girls* in (79) or *the parking* in (103), for example. Usage types of articles will be taken up in the next chapter and we will also explore the interaction between the meaning of a given noun in an NP and the meaning of its accompanying article.

It remains now to ask what can be learned about variation in article use from their collocational patterns. To answer, we will look again at the R-1 collocates of *the*.

6.2 Right and left collocates of *the*

This section describes the R-1 and L-1 collocates of the definite article.

6.2.1 R-1 collocates: nouns

Examining the list of R-1 collocates of *the* (see appendix) one sees an abundance of nouns. In 6.1 above it was suggested that many of these top-ranking nouns are found in all varieties because they usually denote large-scale entities that are relevant to everyday human experience. It is important to remember that these nouns are marked as definite. Some, such as *time* and *world*, are frequently used for they do not only denote real-life entities but also metaphorical ones (Mahlberg, 2005; Stubbs, 2007). Others are frequently used because they have other roles in the discourse. One such noun is

'people'. It belongs to a class of nouns identified by Halliday and Hasan as "a borderline case between a lexical item ... and a grammatical item" (1976: 274).

Called 'general' nouns, their function is to create cohesion that could otherwise be accomplished by the use of a grammatical item such as a pronoun. This argument is illustrated by Halliday and Hasan by the following example:

- (1) The ascent
(2) The climb
I turned to the ascent of the peak. (3) The task is perfectly easy.
(4) The thing
(5) It

(Halliday and Hasan, 1976: 279)

The numbered items are all possible cohesive items, ranked from the most specific to the most general from top to bottom. At (4), just before the reference to *the ascent* is accomplished by the use of a pronoun, we have a general noun, *the thing*. This noun, like *the people*, enables a generalised reference to be made to the antecedent.

Halliday and Hasan state that "a general noun in cohesive function is almost always accompanied by the reference item *the...*" (1976: 275). Although this is not always proven to be the case (Partington, 1998: 91-96), the fact remains that the majority of these nouns occur with the definite article. The class of general nouns is not easy to delimit, however. Partington points out that "general nouns are an open set, with

speakers/writers using nouns cohesively all the time” (1996: 96). However, according to Halliday and Hasan, they consist of “a small set of nouns having generalized reference within the major noun classes, those such as ‘human noun’, ‘place noun’, ‘fact noun’ and the like” (1976: 274). This implies the most superordinate member in a set of hyponymic words. It is argued, however, this formulation does not guarantee the identification of all general nouns (Partington, 1998).⁴⁰ He points out that it is the way a noun is used (i.e. cohesively or otherwise) that determines its status as ‘general’ in Halliday and Hasan’s sense. Partington’s argument notwithstanding, if we look at the list of nouns in the appendix, there is a clear difference between these general nouns and many of the rest of them. For example, in ICE-AUS we find *the thing*, which is a general noun. In the same corpus, we can also find *the Liberals*, clearly a specific noun. Compare these with what we find in ICE-IND. *The thing*, as a general noun, is also found here. However, for specific nouns we find NPs like *the sangha* (a community) or *the Maidan* (the name of a popular cricket ground). Very often these general nouns *are* superordinates while the specific nouns *are* hyponyms (e.g. ‘the members’ – ‘the Liberals’, ‘the group’ – ‘the sangha’, ‘the field’ – ‘the maidan’). Notice, however, the superordinate/hyponym criterion would leave us with too many general nouns to deal with. To solve this problem, in our analysis we must differentiate between ‘general nouns’ and ‘very general nouns’ (Chalker, 1996 in Mahlberg, 2005: 7). We will

⁴⁰ Mahlberg (2005) solves this problem by using frequency as her most important criterion in selecting the nouns to focus on. Since she focuses on a single variety (British English) and her objective is to explicate the supportive roles of these chosen nouns in discourse, this method is seen as successful. In the present study, reliance on a single measure of frequency does not seem to be promising because different varieties rank different words differently. For example, ‘matter’ is found to be the 120th most frequent collocate of *the* in ICE-Ken. If we were to depend on its frequency, then we would have to either include a lot of other words from this variety to reach its 120th position or miss it completely. The latter option would be unfortunate because ‘matter’ is known to be a productive general noun.

continue to use the term ‘general nouns’ although, for all intents and purposes, our interest is in the latter.

For the purposes of our analysis, the general nouns to be analysed are those general nouns identified by Halliday and Hasan. They are the following:

people, person, man, woman, child, boy, girl [human]

creature [non-human animate]

thing, object [inanimate concrete noun]

stuff [inanimate concrete mass]

business, affair, matter [inanimate abstract]

move [action]

place [place]

question, idea [fact]

(Halliday and Hasan, 1976: 274)

To include the plural form of some of these nouns, these were also analysed: *persons, men, women, children, boys, girls, creatures, objects, places, questions* and *ideas*. The remaining words in the above list either do not exist in the plural or carry a different meaning in that form.

Although general nouns have many important roles to play in creating cohesion (see Mahlberg, 2005 and 2003), one of the effects of having too many of them is a high level of generality (see Ringbom, 1998 on the use of ‘general words’ by learners and Carter and McCarthy, 1997 on the role of ‘general words’ in vague language). The

usually unintended high level of generality is in fact a characteristic of interlanguage in SLA (Ringbom, 1998). It is this aspect of general nouns that is the focus of this analysis. In the light of the above discussion, it needs to be asked if there is a high level of generality in the Outer Circle varieties. In relation to this investigation, we need to ask about the nature of their specific nouns as well.

To answer these questions, the most frequently occurring nouns with a frequency of at least five were analysed. These were taken from the top 1000 most frequent R-1 collocates of *the* (with a frequency of at least five). However because the results for the combined totals for spoken and written language for each variety were not very informative, the analysis was made on these two modes of language separately. Tables 6.1 and 6.2 present the findings.

Table 6.1: General and specific nouns among the top 1000 R-1 collocates of *the* (spoken)

	BrE	AusE	NZE	AmE	SgE	IndE	PhilE	KenE
general	13	15	15	4	18	17	14	16
	1.7%	2%	1.9%	1.7%	2.4%	2.2%	1.9%	3%
specific	20	12	11	2	20	13	14	21
	2.5%	1.6%	1.4%	1.1%	2.7%	1.3%	1.9%	3.9%
total nouns	786	764	782	175	755	760	744	538
	100%	100%	100%	100%	100%	100%	100%	100%

Table 6.2: General and specific nouns among the top 1000 R-1 collocates of *the* (written)

	BrE	AusE	NZE	AmE	SgE	IndE	PhlE	KenE
general	12	14	12	11	10	11	13	12
	1.5%	1.8%	1.5%	1.7%	1.3%	1.4%	1.6%	1.5%
specific	16	18	12	9	17	11	23	21
	2.1%	2.4%	1.5%	1.4%	2.2%	1.4%	2.9%	2.7%
total nouns	779	764	767	647	767	766	793	778
	100%	100%	100%	100%	100%	100%	100%	100%

Table 6.1 shows that, for the majority of the varieties, a higher percentage of specific nouns is found in their spoken language. Where the use of *the* is concerned, one could observe that speakers of these varieties incorporate a higher amount of specific information through their use of the ‘the+N’ structure to balance the more general information conveyed by general nouns alone. However, the three Inner Circle varieties of AusE, NZE and AmE seem to defy this trend. It is possible that speakers of these varieties use other kinds of definite NPs to convey specific information.

Table 6.1 also shows that, except for PhlE, there is a slightly higher number of general nouns used in the Outer Circle varieties. It is quite possible then that there is a slightly higher level of generality in the nouns used with the definite article in SgE, IndE and especially KenE. This somewhat parallels the characteristic of ESL learners observed by Ringbom (1998). However we also saw above that speakers of those Outer Circle varieties employ a higher number of specific nouns to offset this large amount of

generality. This is especially true of KenE, which scores the highest percentage of specific nouns.

Table 6.2, on the other hand, gives a rather different description of the varieties. Whereas in spoken language there is a discernible difference between the two groups of varieties in the use of general nouns, the percentages obtained in the Inner and Outer Circle varieties are not found to differ greatly in written language. One could observe that the formality found in many of the written texts has influenced the level of generality in all varieties to the extent that they are similar to each other in this regard. There is a slightly higher percentage of specific nouns in all the varieties except NZE, AmE and IndE. Possibly these three varieties use other kinds of definite NPs to convey specific information. NZE and AmE are found to be consistent in their low use of general and specific nouns, suggesting that their speakers more commonly use *the* with intermediate general nouns (i.e nouns that are neither very general nor specific). PhilE, which patterns with the Inner Circle varieties for spoken language, has a very high percentage of specific nouns in written language, setting it apart from all the other varieties as far as the use of the definite article with specific nouns is concerned. If one were to equate all these nouns with the level of general information, then it could be said that:

- i. spoken language generally contains a higher level of general information marked for definiteness by the definite article;
- ii. except for PhilE, the Outer Circle varieties contain a higher level of general information marked by the definite article in spoken language;

- iii. the written language of the Outer Circle varieties is generally similar to that of the Inner Circle ones in its levels of general and specific information marked by the definite article;
- iv. PhilE contains a very high level of specific information marked by the definite article in written language;
- v. when both spoken and written forms of language are considered, PhilE patterns with the rest of the Outer Circle varieties because of (iv);
- vi. NZE and AmE contain the highest level of intermediate general information marked by the definite article.

Nevertheless we must note that the above analysis is made using type frequency (as opposed to token frequency). This enables a more accurate profile of the variety of nouns used with *the* in the varieties. If we combine this information with token frequency information, as provided in tables 6.3 and 6.4, we have more evidence to support the claims made above. These tables give the frequency profiles of the nouns by providing (i) ten 100-intervals in which the nouns are accordingly ranked and (ii) the number of tokens that occur in each interval. Table 6.3 shows that speakers of Outer Circle varieties do make use of more general nouns. Twenty four nouns in the Outer Circle are found in the first hundred interval as compared to seventeen in the Inner Circle. While there are fourteen general nouns in the second hundred interval in the Inner Circle and only ten in the Outer Circle, the total obtained in the third hundred interval in the latter varieties doubles that obtained in the former, i.e. twelve in the Outer Circle vs. six in the Inner Circle. However notice that the general nouns are spread more widely across the intervals in the Outer Circle, suggesting that some of the

nouns found there are used less frequently than others. This is in contrast to those in the Inner Circle in which there is a greater concentration in the upper intervals, indicating that most nouns are used with high frequencies. Incidentally BrE patterns more with the Outer Circle varieties in this regard (although overall it has a smaller number of nouns than its Outer Circle counterparts). Table 6.4 further illustrates the similarities shared between the two groups of varieties in terms of the use of general nouns in written language. Both groups display similar distributions of these nouns across the intervals.

The behaviour of the specific nouns that collocate with *the* is also of interest. Tables 6.3 and 6.4 show that while general nouns tend to be highly frequently used in the varieties, specific nouns are more variable in their frequencies. This is illustrated by their widespread distribution across the intervals. Of further interest, there is a qualitative difference between the specific nouns obtained in:

- (i) spoken and written language,
- (ii) the Inner Circle data and their Outer Circle counterpart.

Table 6.3: General and specific nouns in 100-intervals (spoken)

	BrE		AusE		NZE		AmE	
	<i>General nouns</i>	<i>Specific nouns</i>	<i>General nouns</i>	<i>Specific nouns</i>	<i>General nouns</i>	<i>Specific nouns</i>	<i>General nouns</i>	<i>Specific nouns</i>
1-100	thing (110) people (97) idea (76) man (44)	British (50)	people (112) thing (56) place (48) person (48) children (49)		people (163) thing (120) person (90) question (80) man (45)	Australians (55)	people (43) thing (39) person (30)	Israelites (16)
101-200	children (38) matter (38) person (31) men (26)	Iraqis (42)	idea (44) matter (36) man (38) women (45) child (37)	ABC (Australian Broadcasting Corporation) (29)	children (41) place (39) idea (39) women (38)	Kiwis (35)	idea (28)	Egyptians (7)
201-300		Americans (23) Romans (19)	men (30) boys (20) woman (23) girls (18)		men (20) matter (18)	marae (a Maori community meeting place) (21)		
301-400		Germans (18) Thames (14)		Commonwealth (10) French (10) Australians (10)	woman (17) girl (15) girls (15)			
401-500	stuff (13)	Tories (13)	stuff (11)	Chinese (6) Christians (6) British (5)	boys (14)			
501-600	women (10)	Kurds (11) pound (11) Israelis (11)				Commonwealth (10)		

601-700	girls (9)	Soviets (10) Abbey (9)				British (9) Aussies (9) pakeha (9) Tasman (9)		
701-800	woman (8)	Exchequer (a treasury arm) (8) Scots (8) Greeks (8) gunners (members of Arsenal football club) (8) Gypsies (8)		Pacific (7)		Americans (8)		
801-900	boys (7)	Gauls (7) Pentagon (7)		Antarctic (5) Aussies (6) Kiwis (6) Jaguar (a car brand name) (6)		Italians (7)		
901-1000		Guardian (6)				waka (a Maori canoe) (6)		

Table 6.3 (continued)

	SgE		IndE		PhiE		KenE	
	<i>General nouns</i>	<i>Specific nouns</i>	<i>General nouns</i>	<i>Specific nouns</i>	<i>General nouns</i>	<i>Specific nouns</i>	<i>General nouns</i>	<i>Specific nouns</i>
1-100	people (160) question (105) thing (102) person (56) place (47) idea (45) man (44)	PAP (People's Action Party) (131) Philippines (67)	people (330) person (55) place (58) thing (51)		people (231) question (88) person (61) thing (54)	Philippines (543)	people (122) women (65) person (62) question (56) woman (57) children (52) man (48) idea (44) men (44)	French (23) UN (21) US (21)
101-200	children (35)	Indonesians (37) British (28)	matter (48) man (39) idea (34)		girl (40) place (37) matter (34) child (30)	Americans (41)	thing (20) place (15)	Kikuyu (an ethnic group) (10) Africans (11) harambee (a community tradition) (10)
201-300	boys (23) woman (24) men (20)	LTA (Land Transport Authority) (24) Malays (20) French (10)	children (30) object (30) child (25) women (22)	British (18)	men (28) children (27) man (24)	Comelec (Commission on Elections) (25) Filipinos (23) Visayas (name of an island group) (20)	boy (14) boys (13)	bible (10) Czechs (9) Luo (an ethnic group and/or its language) (8) British (8) matatu (a bus-like mode of transport) (8)
301-400	move (17) child (17) girls (16) matter (15)	SDP (Singapore Democratic Party) (16) SAF (Singapore Armed Forces) (15)	girls (20) boy (17) boys (17)		women (18)	bible (19) Netherlands (16)	matter (9) girls (9)	Germans (7)

401-500	women (13)	yen (13) Malaysians (12)	girl (11)	Chinese (7)		EDSA (a highway) (15) Spratlys (14)		Tempest (a play title) (6) Beatles (6) kikuyus (members of the Kikuyu group) (6) maasais (members of the Maasai group) (6) Luos (members of the Luo group) (6)
501-600	girl (11)	IMF (International Monetary Fund) (11) durian (11)				peso (12) Cubans (12)	girl (6)	
601-700		ANC (African National Congress) (10) Indians (10) Singaporeans (10) NTUC (National Trades Union Congress) (9)	woman (12)	Aryans (10) Indians (10)	woman (10)			KMC (Kenya Meat Commission) (5) NSSF (National Social Security Fund) (5) DC (Deliverance Church) (5)
701-800		Americans (8) Thais (8)	men (9)	maidan (a field and the name of a popular cricket ground) (8)		cordillera (a mountain range) (6)		Akamba (an ethnic group) (3)
801-900		Commonwealth (7) RODA (Registration of Deeds Act) (7)		sangha (community) (7) French (6) Commonwealth (6)	girls (7)			
901-1000	boys (7)							

Table 6.4: General and specific nouns in 100-intervals (written)

	BrE		AusE		NZE		AmE	
	<i>General nouns</i>	<i>Specific nouns</i>	<i>General nouns</i>	<i>Specific nouns</i>	<i>General nouns</i>	<i>Specific nouns</i>	<i>General nouns</i>	<i>Specific nouns</i>
1-100	people (38) idea (34)	British (32) UK (39) US (34)	people (49) matter (46) person (43) question (40) child (38)	NRMA (National Roads and Motorists' Association) (30)	people (69)	British (29)	people (27) thing (24) person (7)	polka (15)
101-200	matter (23)		idea (28) children (26)	Commonwealth (24)	men (40) man (38) child (38)		idea (5)	Republican (10)
201-300	man (18) woman (17) children (19)	USA (18)	woman (24) man (23) place (17)	Maxima (a car brand name) (22) Cressida (a car brand name) (20) MZ (a car brand name) (18)	person (29) matter (26) place (26)			Americans (8)
301-400	girl (15) men (14) child (14)	Tories (13) Romans (13) Atlantic (13)			woman (22) girls (18) women (19)			IRA (Irish Republican Army) (7) Pentagon (7)
401-500	women (11)	Franks (a Germanic tribe) (12) BBC (12) CIA (11) Tory (10)	move (12)	SSAT (Social Security Appeals Tribunal) (13) Antarctic (11) Pacific (7) ACTU (Autralian Council of Trade Unions) (7)		USA (17) US (17)		US (6)

				ALP (Australian Labor Party) (7)				
501-600	girls (9)		men (9)		boy (13)	Chathams (a group of islands) (14) Wahine (a ferry service) (13)		Olympics (5)
601-700		Falklands (9) EC (9) Americans (9) Italians (8)		aborigines (6) IRC (Industrial Relations Commission) (6) XJR (a car model) (6) Liberals (6) ICAC (Independent Commission against Corruption) (6)	boys (11)			Republicans (4)
701-800			boy (7)			OECD (Organisation for Economic Cooperation and Development) (10) Wallac (a calibrating instrument brand name) (9)		
801-900	boy (6) boys (6)		women (5)	Mack (a lorry/truck brand name) (5) British (5) Turks (5) EC (5)		marae (9) BNZ (Bank of New Zealand) (9)		
901-1000		HSE (Health Service Executive) (6)				UK (7) DSIR (Department of Scientific and Industrial research) (7)		

Table 6.4 (continued)

	SgE		IndE		PhilE		KenE	
	<i>General nouns</i>	<i>Specific nouns</i>	<i>General nouns</i>	<i>Specific nouns</i>	<i>General nouns</i>	<i>Specific nouns</i>	<i>General nouns</i>	<i>Specific nouns</i>
1-100	child (71) people (68) children (38)	PAP (64)	people (82) matter (48) question (33)		people (90) women (39)	Philippines (484) Internet (57) MILF (Moro Islamic Liberation Front)(56) Marcoses (48) peso (28)	people (114) man (66) child (54)	Kipsigis (an ethnic group) (30)
101-200	man (32) person (24) idea (22)	British (26) Malays (23)	place (27) women (23)	USA (18) IAAS (Indian Association for American Studies Trust) (18)	object (33) man (31) child (29) question (29) men (29) person (29) place (28) woman (20)	Americans (25) Katipunan (a revolutionary political organisation) (22) PCGG (Presidential Commission on Good Government) (22)	women (43) girl (33) woman (34) boy (27) matter (21)	KWS (Kenya Wildlife Service) (21)
201-300		bangsawan (a traditional opera) (18) SAF (19)	person (19) child (18) girl (17)	US (17) BJP (Bharatiya Janata Party) (14)	children (18)	IMF (21) BIR (Bureau of Internal revenue) (20) Pacific (14)		British (20) KBC (Kenya Broadcasting Corporation) (20) ANC (African National Congress) (19) Kikuyu (11)
301-400	men (14)	US (16) EDB (Economic Development Board) (16) HDB (Housing Development Board) (15)	children (14)	BCO (Bombay Chamber Orchestra) (12)	boy (15)	Thomasites (16) Upcat (University of Philippines College Administration Test) (15)	men (15) boys (13) person (13)	

401-500	girl (13)	Indians (13) UDMC (name of an outdoor food court) (10) CNB (Central Narcotics Bureau) (12) PPD (Primary Production Department) (11)	boy (13)	VHP (Vishwa Hindu Parishad Party) (9)		Filipinos (12) EDSA (11) NGEC (New General Education Curriculum) (12) Treo (a mobile phone model) (12)		IPK (Islamic Party of Kenya) (11) Marakwet (an ethnic group) (11) KTN (Kenya Television Network) (10)
501-600	women (11)			LIC (Life Insurance Corporation of India) (7)		Napocor (National Power Corporation) (11) IPI (International Press Institute) (10)		ISK (International School of Kenya) (9)
601-700		Hokkiens (8)	woman (9)			NCCA (National Commission for Culture and the Arts) (9) NEA (National Electrification Administration) (8)	girls (8)	Christians (7) USA (7) US (6) Maasai (6) shamba (a plantation by a local method) (5) Orkoiik (an ethnic group) (5) Europeans (5) CNN (Cable News Network) (5) Africans (5)
701-800		CPP (Cambodian People Party) (7) NSS (Nature Society Singapore) (7)				Comelec (7) MSCB (Multisectoral Commission on Budget) (7) NFA (National Food Authority) (7) NSC (National Security Council) (7) NPA (New People's Army) (7) AFP (Armed Forces of the Philippines) (7)		

801-900		CAD (Commercial Affairs Department) (6) SDP (6)		Narmada (name of a river) (6) Robos (a fictional clan) (6)	girl			Luo (4) Hungarians (4)
901-1000	boys (8)			British (5)		DOJ (Department of Justice) (5) trapo (a corrupt politician) (5) barrio (a poor neighbourhood) (4)		

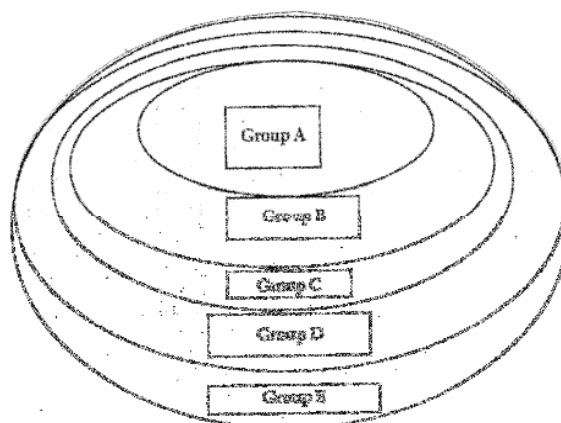
Comparing tables 6.3 and 6.4, one finds a large number of abbreviations that constitute the specific nouns in written language. These are usually acronyms of official institutions. In SgE, however, there are nearly the same numbers of abbreviations in both spoken and written language. This suggests that firstly, the relevant institutions are very significant in the everyday life of SgE speakers (and writers). Secondly, it suggests that the definite article is used with names of institutions very regularly in this variety. On the other hand, for the other varieties the same claims are true for the written language only.

Perhaps unsurprisingly, the Outer Circle varieties tend to use more specific nouns that are locally based, socio-culturally speaking. However, looking at NZE, one sees quite a number of Maori-related words (which will be discussed below). If one compares NZE and AusE, the lack of aboriginal words in the latter antipodean variety is noteworthy. It suggests that there have been different levels of indigenous influence on the variety spoken in each country (see Schneider, 2008).

In the Outer Circle varieties, there is an important difference between the specific nouns used in spoken and written language. The data of written language yield specific nouns that are less local (socio-culturally) than in the spoken data. In fact, there is no qualitative difference between the specific nouns in the Outer Circle's written language and those in the data of Inner Circle varieties, both spoken and written. This claim can be elucidated by Ooi's Concentric Circles Model (1997 in 2001: 111-113), a

lexicographic framework designed to classify lexical entries with ‘nativised’ content (see figure 6.4).

Figure 6.4: Ooi’s Concentric Circles Model (2001)



All examples come from (or at least are used in) SgE. Group A is Core English, which “includes English of most utility worldwide” (p. 111). Examples include ‘kungfu’ and ‘amuck/amok’. Group B includes words of English origin acceptable in formal situations. Some examples are ‘tuition teacher’, ‘red packet’ (money given in a red envelope) and ‘MRT’ (Mass Rapid Transit System). Group C consists of words of non-English origin used in formal situations. These can be exemplified by ‘rambutan’ (a small, hairy fruit) and ‘songkok’ (a type of traditional headgear). Group D includes words of English origin but used in informal situations only. An example is the use of ‘send’, to mean ‘take’, in ‘Let me send you to the airport’. The outermost Circle is Group E, comprising words of non-English origin and used in informal situations only. Some examples are ‘kiasu’ (fearing to lose out) and ‘malu’ (shameful).

It appears from Table 6.4 that the acronyms which form the specific nouns in the written language of the Outer Circle varieties belong in Group B. Some examples are *HDB* (SgE), *BIR* (PhilE), *BJP* (IndE) and *KTN* (KenE). However it must be noted that their possible promotion to Group A is only denied by their limited function, and thus currency, outside their own countries. Compare these nouns with *BBC* and *CIA*, which can be argued to be in the standard English lexicon. Interestingly, there are specific nouns in the written language of the Inner Circle varieties that share this characteristic of limited function/currency. These include *SSAT*, *ACTU*, *IRC* (AusE) and *OECD*, *BNZ* and *DSIR* (NZE).

What sets the specific nouns in the Outer Circle apart from those in the Inner Circle, except NZE, is that in their spoken language there are words that belong in Group C: *maidan* and *sangha* from IndE, *cordillera* from PhilE⁴¹ as well as *matatu* and *harambee* from KenE. (There is none in spoken SgE but one, *bangsawan*, is available from its written data while another one, *trapo*, is available from PhilE written data.) Most of them are found in both formal and informal registers. NZE presents a similar case with its specific nouns *marae*, *waka* and *pakeha*. However, the most important difference between NZE and the Outer Circle varieties is that these NZE nouns have been formally incorporated into the variety with their own entries in at least one dictionary, i.e. the *Encarta World Dictionary* (see Ooi, 2001 for the same effort that is being done for SgE).

⁴¹ Originally Spanish.

We thus see now that in Outer Circle varieties there is a tendency to use the definite article with nouns with local sociocultural content (i.e. when they are definite). If anything, this proves that the noun quality of the indigenous words is preserved when used in English and this in turn allows them to take the definite article to form an NP as required by the grammar of this language.

6.2.2 R-1 collocates: adjectives

The analysis of adjectives among the top 1000 R-1 collocates (see appendix) of *the* reveals that adjectives which best differentiate the varieties (but not necessarily their patterns of use) are also those that are high in socio-cultural content (see specific nouns above). Table 6.5 shows these adjectives, identified from the top 1000 most frequent collocates of *the* in both groups of varieties. Proper nouns used as adjectives are included in this analysis.

Table 6.5: Adjectives with high socio-cultural content among the top 1000 most frequent R-1 collocates of *the*

Variety	Total adjectives	Adjectives with high socio-cultural content	Adjectives
BrE (spoken)	169	28 (16.6%)	Soviet, European, French, Iraqi, American, English, London, England, Israeli, Irish, German, Saudi, Roman, Greek, Arab, Arsenal, Jewish, Japanese, Ferndale, Midland, Kuwaiti, Christian, Russian, Republican, Palestinian, Easter, Italian, Australian
AusE (spoken)	187	24 (12.9%)	Australian, Labor, Soviet, American, Liberal, aboriginal, Olympic, French, Anglican, Japanese, English, Chinese, Melbourne, Queensland, Christmas, Derby, Aussie, Macquarie, Russian, Roman, German, Italian, Jewish, Christian
NZE (spoken)	128	20 (15.6%)	Maori, Australian, Auckland, Clyde, Wellington, kiwi, American, French, Waitangi, Japanese, Italian, Gulf, Caxton, Wanganui, European, Indian, Dunedin, Russian, Porirua, Otago

AmE (spoken)	35	0 (0%)	
SgE (spoken)	128	25 (19.5%)	Singapore, Chinese, Soviet, Thai, American, Japanese, Malay, Vietnamese, ASEAN, Indonesian, French, Asia, Indian, Asian, Malaysian, European, Nanyang (originally a region name), Sabah (a state), Russian, Singaporean, Kedah (a state), Negri (a state), Bosnian, Sarawak (a state), NUS
IndE (spoken)	181	29 (16%)	Indian, Soviet, English, Kashmir, American, Gulf, Maharashtra (a state), Punjab, Pakistan, India, Russian, Kerala (a state), French, Hindi, Pakistani, Hindu, Bombay, Saudi, Tamil, Kannada (a language), Nobel, Mandal (a proper name), European, Jammu (a state), Karnataka (a state), Haryana (a state), Iraqi, Asia, Telugu
PhilE (spoken)	193	27 (14%)	Philippine, Filipino, Chinese, American, Manila, Spanish, Olympic, Estrada, Soviet, Aquino, Asia, Rizal, Greek, Adamson, Batasang (a region), Asian, Marcos, Lasallian, Christian, Batasan (a region), Ninoy, Cebu, Arroyo, ASEAN, Ramos, Indonesian, Subic
KenE (spoken)	143	21 (14.7%)	Kenya, African, Kenyan, French, Nairobi, Portuguese, Soviet, Kiswahili, German, English, Tanzanian, Faro, Bosnian, Christian, Kenyatta (a proper name), European, Nigerian, Haitian, Indian, Baltic, Angolan
BrE (written)	208	20 (9.6%)	Soviet, European, Gulf, American, French, London, English, Scottish, Bloomsbury, anglo, Iraqi, Indian, Italian, Japanese, German, Chinese, Serbian, Jodrell (a proper name), Christian, Yugoslav
AusE (written)	215	25 (11.6%)	Australian, Sydney, Victorian, Liberal, aboriginal, Japanese, American, Indian, German, Medicare, Cornish, Queensland, Adelaide, Akai (a video cassette recorder brand name), European, Hawke, Hawkesbury (a river), Indonesian, English, Cherokee, Alfred (a proper name), Austrian, Remington (a rifle brand name), Arabian, Melbourne
NZE (written)	193	21 (10.1%)	Maori, Auckland, Wellington, Soviet, Kiwi, English, Maui, Awarua, Waimakariri, Chatham, Australian, German, European, Waikato, Hutt, Roman, Marlborough, Victoria, Manganui, Otago, Canterbury
AmE (written)	190	17 (9%)	American, British, Soviet, Japanese, Republican, French, Irish, German, Illinois, European, African, Chinese, Clinton, Atlanta, DeKalb, Russian, Vietnam
SgE (written)	223	12 (5.3%)	Chinese, Singapore, Soviet, Japanese, NUS, Indian, Hainanese, Cantonese, Tao, IGES (Initial Graphics Exchange System), Asia, Asian,
IndE (written)	219	25 (11.4%)	Indian, Gulf, Bombay, Soviet, Vai (a town), English, American, Delhi, French, Kashmir, Japanese, Maharashtra, Mysore (a city and/or state), European, German, Hindu, Pakistani, Shivaji (a proper name), Pakistan, Maruti (a car brand name), Mandal (a proper name), Hindustan, Goa, Andhra, Narmada (a river)
PhilE (written)	196	25 (12.8%)	Phillipine, Filipino, Asian, Marcos, American, Japanese, English, Ramos, Manila, French, Spanish, Chinese, Asia, Estrada, Catholic, Labor, Christian, ASEAN, Aquino, Sulu, Mindanao, Rizal, Maryknoll (Catholic Foreign Mission of America), Malolos (a city), Roman
KenE (written)	210	18 (8.6%)	Kenyan, Nairobi, Swahili, English, Thome (a proper name), Eldoret (a town), Japanese, Somali, American, Kenyatta (a proper name), Kakameza (a region), Bantu, Africa, German, Christian, French, Kerio (a river), Danish

The table shows that very often the adjectives are formed from proper names of persons and places. Many of them received derivational affixes to function as adjectives e.g. *Cantonese*, *Iraqi* and *French*. Because of the preponderance of these proper names, the adjectives easily reflect the local socio-cultural realities in the country in which a given variety is spoken. For example, the adjectives in PhilE clearly show its preoccupation with the country's history and politics. At the same time we find adjectives like

American, Chinese or even *French* in many of the varieties, prompting speculations about the socio-cultural significance that these words represent at the global level. Nevertheless, it is worthwhile to note that it is these highly specific meanings in the adjectives that frequently bring out the definite meaning required for definiteness marking by *the*.

6.2.3 L-1 collocates of *the*

In the first position to the left of *the* (henceforth L-1), a large number of words commonly shared between the varieties were found. The majority of these are prepositions. Figure 6.5 contains a list of the most frequent prepositions found in all varieties. Because of the large number of overlapping words in the L-1 position, the analysis was only based on the top 100 words in all varieties.

Figure 6.5: Most frequent prepositional collocates of *the* at L-1 position in all varieties

1. of	11. into	21. within	31. throughout
2. in	12. during	22. up	32. inside
3. to	13. over	23. out	33. until
4. on	14. after	24. around	34. among
5. for	15. through	25. down	35. while
6. at	16. between	26. toward(s)	36. until
7. by	17. under	27. across	35. whether
8. with	18. like	28. off	
9. from	19. before	29. along	
10. about	20. against	30. outside	

The number and variety of nouns shown in the above figure reflect Biber et al.'s observation that nouns, as compared to pronouns, are very frequently used as

prepositional complements (1999: 235). Data from this figure can be compared with the data in figure 6.8 (top L-1 prepositional collocates of *a/an*) and we find that prepositional complements tend to be definite rather than indefinite. If prepositions are said to typically express “circumstantial meaning” (e.g. Collins, 1998: 32), then many of the referents that constitute this meaning are considered known to the hearer (see chapter 2). One preposition that is found only in the top 100 words of Inner Circle varieties (except in the American data) is *round*.⁴² Concordance 6.1 was generated for *round the*, showing 50 randomly selected lines from the Inner Circle varieties and 50 randomly selected lines from the Outer Circle varieties.⁴³

Concordance 6.1: 50 lines of *round the* from Inner and Outer Circle varieties

	Outer Circle
My tongue wandered	round the little square bracke
t: Quay cranes at TPT operate	round the clock.
nd walked and swerved and ran	round the people, else I woul
photocopied files and painted	round the office.
oming and the end of the year	round the corner.
air traffic controllers work	round the clock to ensure smo
omputer ah break down convert	round the CPU
n fact when I I've been going	round the countries to preach
Trying to actually go	round the Pahang middle instea
He's reading his way	round the defence so easily an
e last eleven years I've been	round the uh army camps the b
ike not knowing what's coming	round the corner
So it's like you're working	round the whole year you know
Space shuttle	round the world
Eh actually we should walk	round the buildings
Hairy stood up, went	round the stack of files, bent
The next day the story went	round the block that shankar h
Trim	round the edges and remove scr
the facility must be utilized	round the clock so as to make
puppeteers from 28 countries	round the world, perhaps the
or the moon to go once	round the earth in twenty-eigh
as won the two words	round the world Yacht race whi
ief operations are continuing	round the clock to provide fo
ief operations are continuing	round the clock in the

⁴² There are 1361 tokens of *round the* in COCA with similar uses shown for the Inner Circle varieties here.

⁴³ Only 57 tokens of *round the* were found in the Outer Circle varieties.

(continued)

A control room is working	round the clock in New Delhi
kind of outline all	round the end of the bedsheet
as far as uh if you go	round the whole Jalandhar you
Lungi was tied	round the neck
inst Rod Laver in the opening	round the opening match of the
e have are going	round the from the state
he would how nice to go	round the world again
positions of the earth	round the sun
earth BB-Work	round the sun
The earth is moving	round the sun and the
It has moved	round the sun
And the movement of the earth	round the sun is
henever he went for functions	round the country. In the las
the moral re-armament fellows	round the bend. Whether or no
sion just 24 hours ago. From	round the country came reports
rom him. Accused then took me	round the stores. I took over
istened. Then I heard. I went	round the house, and you opene
and Baba Pesa was chasing him	round the yard with his shotgu
und that it was the other way	round the girls had dropped ou
the city government must work	round the clock to earn the cr
robi cuisine is probably just	round the corner Your
e there was an election fever	round the corner uh Did they r
cause I see someone just goes	round the car and don't test t
ralia on sugar They have gone	round the country to identify
ointed as a minister you went	round the sugar growing areas
t savings needed by borrowers	round the world.

Inner Circle

the water would work its way	round the washer and the tap w
on man (Lee Majors) half way	round the world, any New Zeala
be a long race, three times	round the one-mile circuit, an
leep of Affco owned buildings	round the perimeter of the roa
oa in the course of a voyage	round the world.
have made his projected path	round the south of New Zealand
The ship sailed	round the Cape of Good Hope, a
They rowd round and	round the ship defying and th
e did some deliberate bending	round the tree so it could sta
people of aotearoa right	round the country
vid impressions of the ladies	round the big table
s yacht in the next whitbread	round the world race
urteen boats in the whitbread	round the world race are
no harmful um leakage	round the door at all
of the microwave is the seal	round the door
told me she'd been wandering	round the village and had seen
The walk	round the lake was nowadays ob
pids that were surely waiting	round the bend.
nsidered it demeaning to pass	round the hat to foreigners on
ts to pretend good times were	round the corner.
high, two and a quarter miles	round the base, with cliffs re
his company is spread thinly	round the UK.
y arts festivals, which go on	round the country particularly
she's been touring	round the country as part of h
advisers and experts half way	round the world and come back
and I've liked to think that	round the corner of supreme ag
elp	Round the corner were The Lila

(continued)

u can queue up and get a tour round the White House
miles or four thousand times round the world
Leaders from all round the world telephoned Dow
the fellow traveller hanging round the fringes.' W2F-004
n't call out, and I walked on round the corner without knowi
t your whole idea of yourself round the fact of being consid
ing walks - one day clockwise round the village & next day a
(B):122 At nights we'd gather round the bar and sing old son
er a hearty dinner and coffee round the fire we all settle f
pected to finish first in the round the world solo yacht rac
rs of the judiciary even with round the clock police bodygua
033(C):48 They had an oval ah Round the oval they had a um a
robably be going um somewhere round the fourteenth of Octobe
but um after all that rapping round the night before at top
11(B):175 But you know I mean round the city you don't need
S1A-009(A):54 I brought one round the other night but I at
-005(B):399 It affects me all round the neck and ears S1A
berry tree growing in a crack round the side S1A-002(A):1
orld too One brand New Jersey round the garbage tips S1A-
I have eaten my way round the Yorkshire Dales
are Just put a bit more soil round the sides there and you'
Sapphire and cubic zirconia round the outside
There's an interesting band round the walls

The preposition *round* takes an NP and usually functions as an adverbial. The above concordance shows that there is a larger variety of nouns that collocate with *the* in the Inner Circle than in the Outer Circle varieties. Whereas in the Outer Circle varieties we find a repetition of nouns such as *clock*, *world* and *sun*, in the Inner Circle ones we find, in addition to these, *washer*, *ship*, *tree*, *door*, *hat*, *walls*, *bar*, *fire*, *night*, *oval* and many other nouns. One reason for these differences can be found through an examination of the verbs that are modified by this PP. In both varieties, many of the verbs occurring with *round the* are dynamic. However there is a crucial difference between them. In the Outer Circle, it is observed that these verbs overwhelmingly occur as processes e.g. *wander*, *operate*, *ran*, *come*, *walk*, *went*, etc. In contrast, some verbs in the Inner Circle come across more as events e.g. *owned*, *is*, *were*, *is spread*, and *gather*. The variety of the verbs in the Inner Circle leads to the phrase's productivity and hence,

the use of the definite article with a greater variety of nouns. Concordance 6.1 also shows that there are more instances of transitive verbs being used with the phrase in the Inner Circle group, e.g. *did, made, owned, get, put* resulting in the use of a higher variety of nouns in the PP than that found in the Outer Circle group.

Next we examine the other parts of speech that are found in the L-1 position. Figure 6.6 shows the other words that are found in the top 100 in all varieties.

Figure 6.6: Most frequent non-prepositional collocates of *the* at L-1 position in all varieties

1. and	11. be	21. what	31. you	41. the	51. says
2. that	12. or	22. now	32. both	42. time	52. until
3. is	13. because	23. which	33. that's	43. called	53. even
4. all	14. know	24. get	34. it	44. using	
5. are	15. so	25. take	35. how	45. got	
6. but	16. said	26. think	36. has	46. it's	
7. was	17. not	27. since	37. given	47. been	
8. when	18. see	28. than	38. do	48. mean	
9. if	19. where	29. only	39. were	49. just	
10. have	20. then	30. use	40. however	50. near	

It is apparent from the above figure that *the* occurs after a variety of parts of speech (including contracted forms of *is* as in *that's* and *it's*). Verbs form the largest group, illustrating the extensive use of (definite) NPs as complement in clauses headed by this lexical class. *Be*-verbs are also represented in this group. The next significant parts of speech are subordinators and coordinators. The adverbs *just* and *only* are included here indicating how closely related they are to the meaning of the definite NP rather than the verb. Two examples illustrate this:

107. Singaporeans separated by *just the age* of a generation could exhibit quite different language habits <ICE-SIN:W1A-008#X110>.
108. This is *just the period* for which original steerage passenger lists have often not survived <ICE-NZ:W2D-019#76>.

The pronouns *you* and *it* form the bulk of the noun group found in the figure. *Time the* appears to defy common rules of phrasal constituency. A closer examination however reveals that it is used in a variety of adjuncts usually as part of a PP, e.g. *by the time*, resulting in the unusual *time the* collocation. Some examples are given below:

109. By the *time the* plants became lethal, the race may have forgotten how to control the machines <ICE-IND:W2F-010#X110>.
110. I keep my eyes on the quay the whole *time the* boat is pulling away <ICE-NZ:W2F-011#124>.

As for *the* itself, it is a frequent L-1 collocates too because of the ubiquity of repetition in spoken language.

The words that are found only in Outer Circle varieties are *also* and *this*. Concordance 6.2 presents ten randomly selected lines of *also the* from the Outer Circle varieties and ten randomly selected lines from the Inner Circle varieties.

Concordance 6.2: 10 lines of *also the* from Inner and Outer Circle varieties

Outer Circle varieties

Science classes also the boys they don't care
lector of Central Excise and also the Collector is required
ublic-Decision Making'. It is also the context in which the
talian setup and uh they have also the French setup And uh w
ories and that kind of thing Also the modern changes have m
ot just that but uh it can be also the physical emotional in
That means even now also the process of industrial
t the civil victims but it is also the promoter of internati
“ You were also the reason I had to final
Also the solar and lunar tides

Inner Circle varieties

to mix the metaphor, they are also the joker in the pack.
retson and Ulf Zimmerman, and also the sprinter Abdoujaparov
discipline based approaches but also the fact that in our scho
d J54 121 Aristocrats were also the men most interested i
he side dining room, which is also the area in which smoking
iling. <ICE-NZ:W1B-020#112:11>Also the report and papers fr
ions for business studies but also the other parts be
on perhaps but there's also the citizens' initiated r
sociologist, Mike Grierson is also the warden of a small blo
velopment of young people and also the needs of our economy

The above concordance shows that there is no substantive variation in the nouns found in the two groups of varieties. What seems to be variable and quite possibly explains the more frequent occurrence of the collocation *also the* in the Outer Circle varieties is the use of *also* as a verb modifier. In the Outer Circle varieties, the adverb is sometimes transposed and appears as a postmodifier, e.g. *have also the French*. This use may have led to the increased incidence of *also the* in these varieties.

An examination of the concordance for *this the* shows that it is frequently the result of repairs, occurring only in conversations and far more frequently in the Outer Circle varieties. Other than repairs, speakers use *this* as a general demonstrative pronoun in phrases such as *due to this, for this, in order to do this* and so on. It was found that speakers of Outer Circle varieties tend to use *this* as a general demonstrative pronoun

more frequently than their Inner Circle counterparts (cf. the use of general nouns by learners in 6.2.1). *This the* is also the result of what can be written as *this, the* which signals apposition as in *we have noted reports of this, the other element* or a preposed object Pro as in the second line of the Inner Circle lines in concordance 6.3. The *this the* pattern, too, occurs more frequently in the Outer Circle varieties. Concordance 6.3 presents ten random lines of *this the* from each group of varieties.

Concordance 6.3: 10 lines of *this the* from Inner and Outer Circle varieties

Inner Circle varieties

mbs in Pacific Ocean, due to this the cold war emerged betw
y. This the EDB undoubtedly recog
For this the effect of these two p
o change the what do you call this the maps the maps once i
sually imagine Akamba is this the most reliable
o uh we have noted reports of this the other element of peop
s a priest So I had looked at this the way they they
o how we could try and reduce this the use of patients to a
And then this the skipper the captain g
en we go on to soil fertility This the last part uh after th

Outer Circle varieties

this is your first attempt of this the question, or is this
ve action. When Ike mentioned this the old man K19 93 con
But is this the whole picture,
):263 Back in New South Wales this the annual car race is a
owledgment S1A-013(A):3 Is this the physics area or somet
're getting S1A-038(A):107 This the professions have not
Uhm, now this the other the other main
he's got to bowl from this the northern end bowling
police constable was this the first er as you descr
In unstable ground like this the alignment of turbine

As for the Inner Circle varieties, the following are the L-1 collocates that are only found in them: *well, put, behind* and *round* (see above). These words, however, were not found in the limited American data.

In the case of *well*, fewer tokens were found in the Outer Circle varieties. In her study of this discourse marker in Xhosa English, de Klerk (2005) similarly found that there were fewer types of meanings of *well* in her corpus than those found in native varieties. Pointing to the fact that *well* has an array of meanings with subtle differences between them and that the equivalent of *well* is not available in Xhosa, she argues that speakers of Xhosa English may not be accustomed to the full range of meanings that this word has (2005: 1119-1200). This seems to be the case of the Outer Circle varieties too. With the fewer occurrences in the Outer Circle data, it is possible that some of the roles of *well* are carried out by other discourse markers, especially in spoken language. SgE, for example, is well known for its wealth of discourse particles variously derived from its substrate languages (Lim, 2007).

In the case of *put*, there are additional meanings found in the Inner Circle varieties which are absent or rare in their Outer Circle counterparts. Concordance 6.4 contains ten selected lines from each of the group of varieties. Notice that the main meaning of the verb i.e. to place something (e.g. *Encarta Online Dictionary*, 2009) is evident in all the lines from the Outer Circle varieties. We also find one instance of *put* used as a light verb in *put the blame* and one instance of *put* with the meaning of 'estimate' in *put the death toll at*. In the lines from the Inner Circle varieties all these meanings are similarly found. In addition to these, however, we find the use of *put* in phrasal verbs *put on* and *put down*. There is also an instance of *put* in an idiom i.e. *put the wind up*. Obviously this verb is more prolific in the Outer Circle varieties, resulting in a higher number of tokens of the collocation *put the*.

Concordance 6.4: 10 selected lines of *put the* from Inner and Outer Circle varieties

Outer Circle varieties

ent the procedure but did not put the antibiotics.
ying is We don't even want to put the blame on the man's sho
The existence of laws has not put the courts or jails out of
s and Croates but police have put the death toll at seventy
iform there is good reason to put the fear of the Lord into
ics have opened a meeting to put the finishing touches to a
you find that it is necessary put the group together start
unity for members of staff to put the interest of those stud
subject did the research and put the pieces together all o
ill the roads be repaired? I put the question to the acting

Inner Circle varieties

nt because they've never even put the criteria down on paper
ge out and just just about to put the garbage out and the ph
ogically, from the moment we put the house on the market, w
he could submit to Hawke or put the issue before the party
The police put the figure at fifteen hund
y. This put the blame on Jamison's dec
ttes S1A-022(A):54 Um I'll put the jug on S1A-022(B):5
time to put the stereo on I think -
ks a lot, mate! You certainly put the wind up and out of me

The collocation *behind the* in the Inner Circle varieties also has a few patterns of use not found in the Outer Circle ones. In both groups of varieties, *behind the* is found to form a PP and functions as a postmodifier to nouns, a predicative complement following a *be*-verb, an adverbial and a complement of another preposition in a PP such as *from behind the tree*. Regarding the last type, there are prepositions which occur with *behind the* in the Inner Circle varieties but are absent in the Outer Circle ones. These are exemplified in concordance 6.5 where it presents selected lines of 'preposition + *behind the*' in the Inner and Outer Circle varieties.

Concordance 6.5: 5 selected lines of 'preposition + *behind the*' in Inner and Outer

Circle varieties

Outer Circle varieties

daughter of Njeri, came from behind the house and stood at
Bermuda shorts appeared from behind the giant palm trees,
indanao that this threat from behind the garden is
the questions which are from behind the book
or and Fatsot peeked out from behind the plants like a pair

Inner Circle varieties

brought them two bottles from behind the counter and then we
this design the tiles go down behind the cabinet instead of
just creep around behind the supermarket
nacker takedown. Then fall in behind the inside boat E06
nd cruise ships were gone. Up behind the harbour the island

As can be seen above, the Outer Circle varieties only use one type of preposition for this construction i.e. *from*. On the other hand, there are five types of prepositions that are used in this way in the Inner Circle varieties and their data show that each is used more than three times. The flexibility of *behind* in its combination with other prepositions in these varieties seems to have broadened its meaning slightly, leading to more occurrences of the *behind the* collocation.

6.2.4 Longer-span collocates of *the*

As discussed at the beginning of this chapter, reliance on particular parts of speech to explain the behaviour of the definite article may not adequately identify and explain the patterns of variation that can be found in between the varieties. The multiple-word, or cluster, level analysis is thus intended to capture longer-span collocates of *the*.

The strength of collocation between the words in a cluster is determined by mutual information, which is calculated by Wordsmith Tools. The lengths of the clusters are of two types: three-word and four-word. These two types of length were seen as a natural progression from the two-word level collocation discussed above. Figure 6.7 shows the commonly shared clusters in the top 200 of all varieties.

Figure 6.7: Clusters found in top 200 in all varieties

one of the	the number of
the end of	the same time
end of the	the case of
part of the	in the first
some of the	in the world
at the end	the use of
the fact that	and in the
at the end of	because of the
out of the	to be the
the rest of	at the same time
this is the	in the past
at the time	members of the
the end of the	in the case
most of the	back to the
rest of the	of the world
is that the	the first time
on the other	in the same
the United States	the other hand
at the same	

At a glance, it seems these clusters offer no useful information since many of them (e.g. *some of the*) do not appear to conform to any rules of constituency that we are used to in typical grammars. However collocations require one to “take into account ...complementation patterns” (Hunston and Francis, 1999: 2). This means we often need to look beyond what is visible in established constituencies. As an example, consider the cluster *and in the*. Traditional analysis of the definite article usually focuses on the NP or the PP. What usually escapes such analysis is the fact the PP ‘in

the + N' is frequently coordinated too. In longer discourse this could serve as a simple coordination as in:

111. Signetics is signing an agreement on Monday with U-Card Incorporated, an NTT subsidiary, to market the cards here and in the region, and manage the card systems <ICE-SIN:W2C-004#83>;

or a coordinator with an adjunct in a complex sentence. e.g:

112. we had no idea of how things were going where to sit or anything, and in the end we were litigants in person <ICE-GB:S2A-027 #131:1-132>.

In addition, *and in the*, when combined with certain nouns, e.g. *end* or *process*, is often found in spoken registers and is used by a speaker in response to his/her interlocutor to mitigate the seriousness of the evaluative value carried by *in the end* alone. Compare the two versions of *in the end* in the following:

113. STACY That is a device that's been available since the early '70s, right? Why hasn't it been readily used? Gen. McCAFFREY Well, there hasn't been an incentive on the part of the agencies to push the regulations through, and I must say that some of our friends in the industry have been resistant to pushing it through. A number of companies have purchased it and put it in the mines themselves. STACY It's a cost factor, perhaps? DAVID BRODER It is not a terrible -- not terribly costly. STACY And in the end who would bear the burden of the cost? Would it be something shared by federal unions or labor unions and/or these mining companies? DAVID BRODER It would be the mine operator's cost. It's a cost of doing business

<COCA:2006/SPOK/CNN_LiveSun>.

114. ...sex in three and a half, four months. SAILOR: Yeah. DR. JACK KEVORKIAN And they went through your therapy. SAILOR: Yeah. DR. JACK KEVORKIAN So what does that say? SAILOR: The fact that they're not having sex now I don't think is the issue. In a funny way, for the first time, while they're not having sex, they're each seeing each other. They are each allowing themselves to be seen. Frequency of sex is absolutely no measure of either sexual satisfaction or marital happiness. DR. JACK KEVORKIAN (voice-over) *And in the end*, John and Keye did not find marital happiness. This week they told us they plan to divorce <COCA:2006/SPOK/NBC_Dateline>.
115. That's right.# for a girl to start, to start worrying about that kind of thing! Is is that a societal pressure? I mean is it a ah, th wi, is there any history of it in the family an and, what did your family, did you family know? #Families don't generally know. And it, it wasn't until many years later that I told anyone at all that's, that's the thing I was saying about bulimia, it's very secret. #*And in the end*, did you, I mean wha, did you get the support of of professionals or or #Yes. Erm, but in the end th o only one had any impact and in general, see seeing a female professional was, I have to say, a lot more helpful than seeing male professionals <BYU-BNC:FL6/S_broascst_discussn>.
- 113a. STACY It's a cost factor, perhaps? DAVID BRODER It is not a terrible -- not terribly costly. STACY *In the end* who would bear the burden of the cost?
- 114a. SAILOR: The fact that they're not having sex now I don't think is the issue. In a funny way, for the first time, while they're not having sex, they're each seeing

each other. They are each allowing themselves to be seen. Frequency of sex is absolutely no measure of either sexual satisfaction or marital happiness. DR. JACK KEVORKIAN (voice-over) *In the end*, John and Keye did not find marital happiness.

115a. I mean is it a ah, th wi, is there any history of it in the family an and, what did your family, did you family know? #Families don't generally know. And it, it wasn't until many years later that I told anyone at all that's, that's the thing I was saying about bulimia, it's very secret. # *In the end*, did you, I mean wha, did you get the support of of professionals or or[.]

Granted the use of other elements (e.g., *perhaps* and *would* in 113) also bring about the desired non-confrontational effect between the speakers. However without *and*, the comments following *in the end* in the alternative versions come across as more confrontational (see Carston, 2002 for a relevance-theoretic perspective).

In figure 6.7, we find a high number of PPs (e.g. *at the end*, *at the end of*, *out of the*, *at the time*, *in the world*, etc.). This observation confirms two things:

- i. the dominance of prepositions in the L-1 position as discussed in the previous section;
- ii. the preponderance of (definite) NPs in the complementation of the PP (cf. Biber et al., 1999: 236).

There is also a large number of NPs in figure 6.7 too (e.g. *one of the*, *the end of*, *end of the*, *members of the*, *the first time*, *the other hand*, etc.). *The United States* appears to be

the only NP with an unambiguous referent. Collectively, the number and variety of clusters shown in the figure is evidence that the use of *the* has often attained idiomatic usage as in *(on) the other hand*. At the same time, the list is also evidence of the fact that for some other clusters, their phrase-like togetherness has also already attained some degree of idiomaticity. Compare, for example, *at the same time* with **at the same hour* and *(the) rest of the* and **(the) rest in the*. It is unsurprising then that these clusters are found frequently in all varieties.

If we take these clusters as an index of the phraseological behaviour of *the*, then it would be interesting to discover which clusters are found in certain varieties or in one particular variety only. Table 6.6 presents the clusters that have been identified according to this criterion. A grouping is only valid when there are at least five clusters unique to it. This is because there are far too many clusters or combinations of clusters that occur only once in a particular grouping. If a grouping has at least five clusters, we will regard that grouping as non-random. Where a cluster is not found in the American data, this is indicated accordingly.⁴⁴

Table 6.6: Shared clusters according to Inner and Outer Circle varieties

In two or more native varieties only	In two or more nativised varieties only	In one native variety only	In one nativised variety only
All the sort of (except US)	none	BrE only the United Kingdom the poll tax	IndE only the help of award for the

⁴⁴ When a cluster is found only in ICE-GB, ICE-Aus and ICE-NZ, there is a possibility that it is a variation pattern unique to this grouping. However, since there is no comparable ICE-US to confirm or disconfirm it, such a claim will not be made.

<p>the beginning of away from the at the top (except US) the size of at the beginning any of the of the things (except US) of the way the point of</p> <p>AusE and NZE only that was the out in the the things that the bottom of lot of the the member for in the south</p>		<p>edge of the the far side at the back the start of the Middle East of the British the way in on the far the secretary of on the one do with the start of the the problems of the Gulf war whole of the</p> <p>AusE only the federal govern- ment the whole of the evidence of be in the parts of the relation to the the side of on the first the type of out to the</p> <p>NZE only the New Zealand the national party the south island the north island New Zealand the the All Blacks down to the of the way the way to to have the out on the cent of the of the field the front of of the New Zealand in the new the twenty two to the end up on the was on the in the back the middle of the</p> <p>AmE only (based on six FROWN files</p>	<p>the award for in the form with the help with the help of for the first time in the state if the Indian for the best of the earth fact that the the state government far as the at the background the field of the chief minister there in the the government of the area of the main points along with the the award for the over the world the right side the people in is the first basis of the all over the world of the accused the Janata Dal to the people increase in the they are the in the present of the public to meet the the side of the life of is on the for the purpose</p> <p>SgE only the Singapore broad- casting the Straits Times of the Singapore the other one in the region of the population the thing is from the Singapore to use the to the other found in the the last time in the market</p>
---	--	---	--

		<p>only) in the United States the New York only in the much of the the effects of structure of the of the American in the US the structure of in the second to the editor the face of the cold war the patenting of the concept of for example the in the face the context of the first movement the major groove the study of the power of in the face of the absence of known as the close to the the state of of the major in the late has been the of the game of the U to the point of the time the civil war of the best director of the change in the the direction of the work of the world and death of the the way to control of the at the university of to the same the sense of the culture of changes in the than in the role in the in the context as in the of the year</p>	<p>the sense that at the national in the sense that and all the the whole thing the pick up from the Singapore broadcasting in the long in Singapore the the type of of the main the next lap to do the of the Chinese the needs of and the arts together with the the way of the national university</p> <p>PhilE only in the Philippines of the Philippines rest of the go to the what is the the Philippines and the V F of the Philippines here in the the Philippines is the bureau of to the Philippines university of the of the senate the reason why of the nation the Filipino people of the Filipino of the department the committee on but in the the Philippine stock member of the government and the the presence of the armed forces the Philippines in the members of the the private sector of the committee Philippines and the in the Philippines the</p>
--	--	--	--

		the top 1	<p>for the Philippines in the meantime in the united the so called</p> <p>KenE only the assistant minister the minister for the hon member that the government Speaker sir the the language of Mr Speaker sir the the process of in the office the government to to the police majority of the that is the end office of the president is the end in the process is the end of to know the per cent of the in the society the trial court the government is of the Kenya head of the role of the come to the the rural areas minister of state the teacher should the high court behalf of the led to the the benefit of the needs of is the one the 2nd accused of the news the minister of the person who meaning of the at the trial who is the end of the news the country and of the land for the government</p>
--	--	-----------	--

The table is quite revealing in several ways. First, we find that there are more unique clusters in the Outer Circle group. Interestingly, this does not turn the Outer Circle varieties into a homogeneous group. This happens to be the outcome of the Inner Circle varieties, at least where commonly shared clusters are concerned. In contrast to the Outer Circle varieties which have none, the Inner Circle varieties share ten clusters between at least three of them. In addition, AusE and NZE share enough clusters between them to form a sub-group by themselves. On the contrary, this sub-grouping is not achieved by any of the Outer Circle varieties. Based on this set of findings, we can suggest that there is a far more stable use of the definite article in the Inner Circle varieties than in their Outer Circle counterparts. The next two sections will examine this possibility.

6.2.4.1 Clusters of *the* in the Inner Circle varieties

What motivates the increased use of the ten clusters in the Inner Circle varieties but not the Outer Circle ones? Let's start with *the sort of*. Upon inspection, it was found that both groups utilise the cluster (and usually, the resulting NP) similarly as in:

116. That's the sort of thing that I I'm really the the the at the very basic materials end essentially doing what we think is the strategic research that in ten or twenty years' time will hopefully be uh bear fruit in that uh in that research <ICE- GB:S1A-088 #74:1>.

117. And I discovered and I in fact told one of them <quote> sir you are an M Phil and is this the sort of English you [,] your [,] yourself writing [,] <ICE- IND:S1A-078#45:1:A>.

However it was discovered that speakers of Inner Circle varieties also form the cluster by modifying the referent of a definite NP with *sort of* – which may also double as a discourse marker – while their Outer Circle counterparts do not, as in:

118. they were noisy er but they were controlled and the police were controlled because there were people within the hall attempting to [,] to er <laughs>i suppose if anything</laughs> antagonise them to[,] to exacerbate the er the process in itself and the [...] all i can remember is the sort of rumbling roar that came from people beating on the windows in the middle of jim bolger's speech er jim bolger's <?>rally</?> <ICE-NZ:S1B-25#34>.

As for *the beginning of*, Inner Circle speakers use this cluster with a wider variety of prepositions than that by those of Outer Circle varieties. These include *at the beginning of*, *in the beginning of*, *from the beginning of* and *with the beginning of*. In the Outer Circle varieties, only *in* and to a much lesser extent, *at*, are used. Additionally, Inner Circle speakers utilise the cluster as part of the direct object of the verbs *mark* and *signal* in sentences like:

119. That International Women's Day demonstration marked the beginning of the Russian Revolution of nineteen seventeen <ICE-AUS: S2B-044(A):45>.
120. as you probably all know one of the problems about loading up the clyde dam is that um there are big slides being recognised <?>in the park</?> as we fill the water up little earthquakes may occur which actually signal the beginning of a big slide <ICE-NZ:S2A-024#60-61>.

Away from the presents an interesting case. The frequencies of the cluster in both groups of varieties are not highly different (0.16 per thousand words in the Inner Circle and 0.119 per thousand words in the Outer Circle). The function of this cluster as a PP adjunct to verbs and nouns is found in both groups. However, in the case of the Outer Circle varieties, only two of them – SgE and PhilE – have it in their top 200. They each contribute a large number of this cluster to the total, nearly matching that of the Inner Circle varieties. While this finding shows that the two varieties pattern more closely with their Inner Circle counterparts, it also reinforces the lack of homogeneity in the Outer Circle varieties.

At the top and *the size of* are two clusters that are similarly used in both groups of varieties but occur in a higher proportion in the Inner Circle group than the Outer Circle one (0.112 per thousand words in Inner Circle and 0.072 per thousand words in Outer Circle). The reason for their preference by Inner Circle speakers is difficult to ascertain based on the data. However it can be speculated that, based on their widespread use in spoken registers, it has to do with their association with informality (but see also the discussion of *the top of* in 6.1.1).

As for *at the beginning* it was found that not only are Inner Circle speakers twice as likely to use of this PP cluster as Outer Circle speakers, but they are also 30% more likely to complement it with the preposition than speakers of Outer Circle varieties.

Of the things is frequently prefaced by *one* in both groups of varieties. However in the Inner Circle varieties, there is a higher tendency for it to be used as part of a larger NP containing the personal pronoun *I* as in:

121. *One of the things I* noticed about organisational behaviour is it was somewhat of an attempt to to become multi-disciplined in the sense that the people in my department came out of sociology or psychology or a few had a business background and we had a couple of deviation partners <ICE-AUS: S1A-038(A):4>.

In fact, overall, *of the things* also occurs before personal pronouns more frequently in the Inner Circle than Outer Circle varieties in sentences such as:

122. *one of the things you* give advice to people is how to beat the system <ICE-NZ:S1B-026#179>.

Of the way is used in the Inner Circle varieties in slightly more ways than those in the Outer Circle ones. First it is frequently fronted by *because* making the combination a common adverbial in the Inner Circle varieties. Second, it also occurs with the PP *in terms of* and third, it favours personal pronouns more than it does in the Outer Circle varieties. Some examples are:

123. I have to confess I um but from the readings I sort of found that because of the way their whole was structured *because of the way* their whole political structure's organised it was impossible for the students to contact those who were actually making the decisions 'cos that was there was supposedly not even

the real um<ICE-AUS:S1B-014(B):52>.

124. Yeah it's very good *in terms of the way* it stores data I mean which is maybe is quite important element.... <ICE-GB:S1A-029 #6>.

125. I think that implausible both because *of the way you* described for me the institutional arrangements between NAB and BNZ, and because of the way that Australian banks in general have valued their archives and supported the cultivation of historical understanding in Australia. <ICE-NZ:W1B-021#112>.

There are no patterns resembling those of (124) and (125) found in the Outer Circle varieties.

The point of is frequently followed by *view* in both groups of varieties. In the Outer Circle varieties, it is also frequently followed by *order*. Other than that, the Inner Circle speakers also tend to use this cluster to introduce another definite NP complement, e.g.:

126. Well there's twenty metres from the *point of the* impact where you hit the other car to the stop line on this on the impact side of the road right and then from that line where the intersection begins to where the intersection ends what would that be <ICE-AUS: S1B-064(A)>.

Overall, it can be said that the ten clusters are used in more distinct ways in the Inner Circle varieties than the Outer Circle varieties. What does this say about the use of *the* in both varieties? First, the determinative is shown to be used in more stable ways in the Inner Circle varieties. This claim, however does not equate stability with a smaller number of collocations used repeatedly by their speakers. Rather, we witnessed this

uniformity in the larger overlap of clusters in at least three native varieties. We also found the evidence in the fact that the Inner Circle varieties tend to use the same clusters no matter how variable the wider patterns turned out to be. The Outer Circle varieties, on the other hand, exhibit a higher degree of creativity among themselves even to the extent being dissimilar to one another. This is not to say that the ten clusters are not found in them. Often the differences between the two groups of varieties are only a matter of proportion. However, the fact that they are not used as frequently in the Outer Circle varieties suggests that their speakers may use other constructions and create other patterns to say what they mean.

6.2.4.2 Clusters of *the* in the antipodean varieties

Given the above generalisations, what can be said about the clusters that are unique to AusE and NZE only? It turns out that not all the clusters are easily distinguished from their counterparts in the Outer Circle or BrE (with AmE being mostly problematic to compare). These include *that was the* and *the bottom of*. The difference appears to largely lie in frequency. Nevertheless the remaining clusters do exhibit a few traits that are predominant only in the antipodean varieties.

Out in the is used almost equally frequently in AusE/NZE and the Outer Circle group. The concordances show that the Inner Circle varieties in the southern hemisphere use this cluster with a large number of referents that denote a confined space such as *kitchen, bedroom property, beehive, car* and so on. Further inspection reveals that *out* is often used as part of a ‘phrasal verb or a verb + preposition’ structure, such as in:

127. Julie and Anthony and I were sitting out in the car and Anthony goes well I'm gonna go and see this band <ICE-AUS:S1A-077(A)>.
128. the incentive effects of having abatement going on for a very large long way doesn't matter so much so you might have a lower a lower abatement rate that is the two of those have to be worked out in the context of deciding whether or not you'd want to have your health premium <ICE-NZ:S2A-050#117>.

Phrasal verbs are found in the Outer Circle varieties too but, in contrast, their meanings usually require NPs whose referents denote open and less restricted spaces. An example is:

129. In the industrial town of Kota violent mobs came out in the streets and shouted slogans <ICE-IND:S2B-020#93>.

AusE and NZE make use of the cluster in such a way too but do so far less frequently. The clearest evidence comes from the more frequent use of the idiomatic *out in the open* in the Outer Circle varieties than that in the antipodean ones. BrE and AmE (albeit based on its reduced comparable data) are more similar to the Outer Circle varieties in their usage patterns of this cluster. Given their relatively high collective frequency, it is possible that in AusE and NZE the increased use of phrasal verbs (or phrasal-like verbs) incorporating *out* has contributed to the frequency of the cluster *out in the* as observed here.

As regards the cluster *the things that*, all varieties tend to use it with *all of*, *some of*, *one of* and *many of* as premodifiers. All varieties too tend to complement the cluster with a

personal pronoun such as *I, we, you*, etc. But there remains a difference. In addition to having more clusters of this type, AusE and NZE also have a higher tendency use *the things that* with verbs when combined with one of the premodifiers listed above, e.g:

130. One of *the things that distinguishes* (sic) the New Zealand debate is the recognition to be given to the Treaty of Waitangi in any changes in our formal constitution <ICE-NZ:W2A-011#13>.

132. I th- I think Uh one of *the things that concerns* (sic) me is not BREAK IN RECORDING here uh if we don't have to <ICE-AUS:s1b-079(A)>.

Such use is found only to a lesser extent in BrE (but not in AmE), and to an even lesser extent in the Outer Circle varieties.

Lot of the is a top-ranking cluster in AusE and NZE. Its behaviour is more regular in these two Inner Circle varieties than in the Outer Circle ones. All except two in the AusE/NZE data (i.e. 97%) are tokens of *a lot of the*. The remaining two are *the lot of the*. In contrast, in the Outer Circle varieties there are 74% of *a lot of the* and 16% of *the lot of the*. One example of *the lot of the* from AusE and one example of the same from the nativised varieties are shown here:

133. Fraser introduced a series of major administrative reforms aimed at improving *the lot of the individual citizen* 'against' the bureaucracy: Freedom of Information legislation, a Federal Ombudsman, an Administrative Review Tribunal and an Administrative Court were all created <ICE-AUS:w2a-012(noone)>.

134. *The lot of the Singapore child* is far better. Recognised as the only resource this

country has, he is well fed, housed, educated and generally well looked after
<ICE- SIN:W2E-002#60-61>.

The variation discovered here emphasises the fact that *(a) lot of the* is a preferred cluster in AusE and NZE when compared to the nativised varieties since *the lot of the* has a different meaning as shown in (133) and (134). In this case, BrE and AmE pattern with AusE and NZE although their frequencies are far lower. The use of *a lot of the* as a non-numerical quantifier will be discussed in the next section.

The last two clusters demonstrate the influence of socio-cultural factors on collocations. This influence, which we also saw in the discussion of R-1 and L-1 collocates above, will be increasingly apparent in the following discussion of individual varieties. The use of *the member for* and *in the south* is more frequent in the antipodean varieties quite simply because their socio-cultural contexts require referents that are related to the realities that can be verbally realised by these clusters, e.g.:

135. look all the trumpet playing in the world all the doublebreasted dark suits as worn by *the member for tauranga* none of these things are going to save the prime minister <ICE-NZ:S1B-059#27>.
136. Now Mr Speaker *the Member for Bennelong* was interjecting earlier <ICE-AUS: S1B-060(A)>.
137. did two thousand maoris living in *the south island* in eighteen forty suffer as much as the tainui people and goodness knows how many of them there were er around the waikato region <ICE-NZ:S2A-049#24>.
138. The DPP made its decision on Wednesday after hearing evidence *in the South*

Australia Supreme Court that Mr Wagner, 69, suffered a heart attack last month and a trial might kill him <ICE-AUS:W2C-007(noone)>.

There are instances of *the member for* in BrE too (but not AmE) but it is not ranked in the top 200 of ICE-GB. (135) and (136) are taken from parliamentary discourse but it is interesting that in the Outer Circle varieties where such discourse can also be found, the cluster turns up only once. This implies that speakers of Outer Circle varieties use a different phrase in referring to their members of parliament (although this does not obtain in the Philippines). (137) and (138) are just two instances in which *in the south* turns out to be proper names or the name of entities particular to the Australasian region. *The South Pacific* is another frequent phrase utilising this cluster (which in fact occurs in BrE and some of the Outer Circle varieties all too rarely).

In the light of these findings, it should be pointed out that the Australian and New Zealand grouping does not produce a very distinct set of patterns that are uniquely antipodean, although the idea did appear to be intuitively appealing. There are a few reasons for this outcome, the main of which is the strong adherence which these two varieties have demonstrated earlier to the Inner Circle group that they belong to (most notably with BrE). Recall that many of the clusters are also found in BrE, although in lower frequencies and/or lower ranking. Having said this, the main factors that contribute to the unique characterisation of the top-ranking clusters in the antipodean varieties are socio-cultural in nature. Indeed, we will see that such factors will very largely characterise the use of unique clusters in individual varieties.

6.2.4.3 Clusters of *the* in the individual varieties

Moving on to clusters that are found in individual varieties, we thus find the increased use of clusters with country names in them, e.g. *the United Kingdom, the New Zealand, of the American, the Singapore broadcasting, in the Philippines, of the Kenya* and so on. Other than these, we find locally based referents such as *the federal government* (for Australia), *of the British* (Britain), *the All Blacks* (New Zealand), *the New York* (United States), *the Janata Dal* (India), *the Straits Times* (Singapore), *of the Filipino* (Philippines), etc. Other less nationally based referents also abound given the larger socio-cultural context from which the data were culled. We see these in *the Gulf War, the Middle East* in ICE-GB, *the national university* in ICE-Sin, *Mr Speaker Sir the* in ICE-Ken and so on. The first two are obviously referents to the (first) Gulf War which took place in the decade of data production. *The national university* turns out to be the name of a major public university in Singapore with which many of the speakers in ICE-SIN had an association while the address to *Mr Speaker* is abundant in the parliamentary register found in ICE-Ken.

Other than these, we again discover that there are fewer distinct patterns in the use of the unique clusters in the individual varieties. These patterns are the result of a little more than idiosyncratic preferences that translate into different proportions in different varieties. This is not to say that this cluster analysis has failed to uncover any substantive variation patterns in the varieties. We have discovered discernible differences between the varieties although we have to say that they are, without the aid of corpus technology, not visible to the naked eye. Even more important, the

homogeneity displayed by the Inner Circle varieties as a whole outweighs the variation patterns that arise from the regional and historical circumstances of AusE and NZE. In other words, they pattern more closely with the rest of the varieties in the Inner Circle rather than develop patterns of use that are unique to them. This is not at all the situation for the Outer Circle varieties. They display a high degree of creativity in their patterns but at the same time their independent developments come at the expense of homogeneity. What does this mean for the variation in the Outer Circle varieties that is central to this study? This, in fact, means something very important – variation in the Outer Circle varieties is very complex. It is *not* as monolithic as the group that they have been portrayed to be.

Next we will examine the indefinite article *a/an*. We will draw further conclusions for the collocational patterns of *the* at the end of the chapter.

6.3 The collocates of *a/an*

This section presents the R-1 and L-1 collocates of the indefinite article, as well as the clusters that it is part of.

6.3.1 R-1 collocates of *a/an*

The collocates of *a/an* at the R-1 position are more variable than those of *the*. There is a larger type/token ratio for the parts of speech. For this reason, comparatively fewer overlapping collocates between the varieties will be displayed in a single chart. Figure 6.8 shows the top 20, top 50 and top 100 R-1 collocates of *a/an* in all varieties.

Figure 6.8: Most frequent L-1 collocates of *a/an* in all varieties

Top 20	Top 50	Top 100
1. very	1. hundred	1. matter
2. lot	2. certain	2. week
3. bit	3. single	3. hour
4. few	4. small	4. more
5. good	5. large	5. couple
6. little	6. way	6. particular
7. new	7. great	7. second
8. long	8. result	8. important
	9. whole	9. woman
	10. man	10. better
	11. big	11. group
	12. time	12. person
	13. problem	13. special
	14. day (except US)	14. major
		15. chance

It is important to remember that these rankings emphasise the overlap between all varieties. It masks a few noteworthy characteristics of these words, e.g., that *couple* is ranked in the top 20 in the Inner Circle varieties alone or that *second* is in the top 50 in the same group of varieties. However the focus at this juncture is on the similarities between all varieties. There remain a few interesting observations that can be made about the rankings. In the top 20, we see that many of the words are placed there due to the indefinite article's property as a numeral. Thus we find a few partitive nouns (Quirk et al., 1985: 251) whose collocations with *a* result in quantifying expressions such as *a lot, a bit, a few*. *A little* is also a quantifying expression although not formed with a partitive noun. *Good* is a productive collocate not only because it is a common adjective but also because it serves as a quantifier, e.g.,

139. But uh <,> in the department itself there was a good deal of discussion <,> about Indian writing in English <ICE-IND:S2A-037#17>.

Great and *couple* are adjective-looking words that form qualifiers and/or partitives, e.g:

140. It's very very useful es especially for the engineering students because they depend a great deal on conferences and uhm symposiums it is unsurprising to find that a number of these words are <ICE-SIN:S2A-051#61>.

141. The subject you have chosen is vast and to do it properly will require a great deal of research on your part <ICE-NZ:W1B-030#28>.

142. We need a couple of months of sustained rainfall to make any difference, NSW Bushfire Service official Ross Smith said yesterday <ICE-NZ:W2C-016(noone):48>.

Interestingly, as in the case of the definite article some of the adjectives that are most commonly used with *a/an* have the potential to be definite, i.e. they too lend themselves easily to definiteness. In fact in 6.1.1 it was argued that it was this tendency to be definite that attracted the definite article to them. Consider *important*, *major* and *special* as examples. What, then, differentiates between the ‘a(n) important/major/special + N’ and ‘the important/major/special + N’ patterns? It seems that with the indefinite article these adjectives are commonly used to ‘classify’ (see Biber et al., 1999 and chapter 2) as in:

143. He was *an important landowner*, a man of high status who also exercised particular judicial functions within the plebs under his control <ICE-GB:W1A-003 #58>.

In (143) we can see that, through the use of the indefinite article, the quality of being important is ascribed to the landowner. This assertion is made clearer by the appositive *a man of high status*. If the landowner were to be described as *the important landowner*, then the NP would not function as a classification, but rather an identification, much to the detriment of the apposition (cf. *?He was the important landowner, a man of high status...*).

Some of the nouns in the figure 6.8 are time periods, i.e. *week*, *hour*, *day* and *second* (although the last one can also function as an adjective), reflecting both the significance of time to human experience and the historically derived function of *a/an* as numeral used for measurement.

As in the case of *the*, differences that apply neatly to either the Inner Circle or Outer Circle group are hard to find. *While* tends to be found more frequently in the Inner Circle varieties. Concordance 6.6 shows 10 selected lines of *a while* from the Outer Circle varieties and 10 selected lines from the Inner Circle ones.

Concordance 6.6 10 lines of *a while* from Outer and Inner Circle varieties

Outer Circle varieties

stop talking to machines for a while and interact with a re
ze my time as you were saying a while ago speaker A laug
een in the pipeline for quite a while and the only way I fee
a blessing, and after a while, I was Mameng the gran
And every once in a while, showers of water fell
Or you say that once in a while we eat it's okay but h
hose who liked it then after a while those who kept quiet a
And in a while news on sport takes a
uch that uhm let me pause for a while ha kasi baka n
that made him wonder for a while as to whether t

Inner Circle varieties

haven't scribbled to you for a while.
was going to say it would be a while since you've made scon
it might be a while since you were a stude
It has been a while since we spoke,
ughs oh it's still taking a while though for all the
B):110 So it's going to take a while to develop appropriate
W1B-004(noone):52 It takes a while to settle back into ro
30 The jobs are going to take a while to come but at least t
to get relief which I did for a while when the padding was t
o doesn't cook except once in a while when she's E14 32 h

The concordance of *a while* reveals a few patterns that are more frequent in the Inner Circle varieties. Concordance 6.6 shows that *a while* is commonly used in phrases such as *for a while*, *a while ago*, *after a while*, *in a while* and *once in a while*. These patterns occur substantially in both groups of varieties. However, in addition to these, it was found that speakers of the Inner Circle varieties also frequently use *a while* to refer to 'a long period of time' by using verbs such as *take*. This usage occurs in the Outer Circle all too rarely. Furthermore, Inner Circle speakers tend to provide a subordinate

clause introduced by *to* following the main clause consisting of *take a while*, e.g. *it takes a while to settle back*. The Outer Circle data show none of this pattern. Finally in the Inner Circle varieties, speakers also use *a while* to be followed by (i) *when* functioning as an adverb, e.g. *which I did for a while when the padding was* and (ii) *since* as a preposition, e.g. *been a while since we spoke*. The last two patterns are absent in the Outer Circle data.

Nice and *moment* are two words that are found in all varieties except KenE. We will investigate *nice* first, followed by *moment*. Concordance 6.7 shows all the 11 tokens of *a nice* in ICE-Ken as well as 11 selected lines of *a nice* from the rest of the varieties.

Concordance 6.7: 11 lines of *a nice* in ICE-Ken and 11 selected lines of *a nice* from all other varieties

ICE-Ken

. May God bless you and have a nice year, month, day and ti
that you had a nice journey from to
Keep away from trouble and be a nice lady and also take care
go back. Always yours Have a nice day, won't you!!! W1
ever meet her nowadays. Have a nice Day! W1B-SK46 Dear
s so happy very happy to read a nice from yo
ltivating across the slope is a nice way towards soil conser
lone until tomorrow.' 'She's a nice woman, that one,' he ch
like a mother to me. She was a nice lady, very devoted to m
y why uh That guy's such a nice guy The Then the ba
just worried I hear there is a nice system there There is a

Other varieties

I I think that this is quite a nice diagram and I can say t
It's rather a nice parrot
018(C):200 No but that that's a nice thing to be S1A-018(
was a warm old man and it was a nice experience. Everywhere
97 pressed they respond with a nice tactile click. The Mode
He seems a nice enough lad.
S1A-023(B):161 I'll get you a nice pressie from Thailand
73(?):416 Five dollars seemed a nice round figure to me s
Actually this time it was a nice candidate

(continued)

H):152 Um the April issue was a nice month over er on the s-
Yeah it 's such a nice feeling right when you

It is clear from the concordance above that there is a smaller variety of patterns of use in ICE-Ken than that found in the other varieties. It is shown that whereas KenE speakers make use of *a nice* overwhelmingly in 'have + a nice' and 'be + a nice' patterns, speakers of other varieties use it as part of (i) a predicative complement for *seem* (ii) an object of the preposition *with* and (iii) a direct object following an indirect object, e.g. *get you a nice a pressie*. Even in the use of 'be + nice', these speakers quite frequently modify *nice* with either *quite* or *rather*.

Concordance 6.8 shows all the seven lines of *a moment* found in ICE-Ken and seven selected lines of *a moment* from the other varieties.

Concordance 6.8: 7 lines of *a moment* in ICE-Ken and 7 selected lines of
a moment from all other varieties

ICE-Ken

n of the car, hoping to steal a moment and do the talking. B
", yet if we were to stop for a moment and reflect more clos
could see you Philip just for a moment. I'm sorry Mama Mutin
nute just a minute Just a moment! We had agreed earlie
ough even uh just just uh a moment There is also the que
ing gown.' requested Philip. A moment later Mr Muasya and P
BET Chairman Dr took a moment to explain why they d

Other varieties

yne, and I'll come to that in a moment, but My Lord the evid
Creativity is not possible in a moment of personal sorrow or
ed meter flat stretch in just a moment, we're uh, they'll ge
applause and as Ray mentioned a moment ago Mr Keating and ah
ckspot. E02 85 Imagine for a moment now that instead of r
ember 1989). Which is not for a moment to deny the pleasures
you might want to take a moment to look through that

It is shown above that KenE speakers use *a moment* as a complement of the preposition *for*. The adverb *just* is also used with *a moment* to emphasise the short period of time it denotes. Other patterns include appearance with *later* as postmodifier and taking up the role of object of the verbs *take* and *steal*. In the other varieties, in addition to these, we find the use of *in* in *in a moment*. With this preposition, *a moment* itself can be postmodified by another PP introduced by *of* e.g. *in a moment of personal sorrow*. The adverb *just* is also used in the *in a moment* PP. Verbs that take a moment as object include *take* and *mention*. *Steal* appears unique to the KenE data.

6.3.1.1 R-1 collocates of *a/an*: general nouns

Although Halliday and Hasan claim that general nouns occur with the definite article (see 6.2.1), Partington has proved that this is not always the case (1998: 92-96). Using ‘move’ as an example, he shows that this general noun also frequently occurs with *a*. What is of interest here is the role of the indefinite article in terms of the kind of cohesion that Halliday and Hasan have identified (see 6.2.1). In her analysis of ‘man’ and ‘woman’ Mahlberg, too, found that these nouns occur most frequently with the indefinite article together with a modifier (2005: 106-107). She discovers that the ‘*a/an* + modifier + man/woman’ or ‘*a* + man + modifier’ NP can also function to create cohesion by establishing some kind of textual relation with an already established meaning. However this is accomplished by the NP by presenting itself as a non-referential entity (ibid.). (143), given above, is repeated below as an example:

143. He was an important landowner, a man of high status who also exercised particular judicial functions within the plebs under his control <ICE-GB:W1A-003 #58:1>.

Table 6.7 shows the token frequencies of the general nouns identified by Halliday and Hasan (1976: 274) that are found to occur with the indefinite article in the R-1 position. These are taken from each variety's top 500 words only. Unlike what has been done for the definite article, the numbers of tokens of these nouns are given in the table without the information regarding their positions in 100-intervals. This modification was necessitated by two conditions. First, there are not enough word types with a token frequency of at least ten to be ranked down to the 1000th position. This is in addition to the fact that the totals in the varieties vary widely, with a range between ± 250 and ± 650 . Second, unlike in the case of *the*, the general nouns that we are interested in are high frequency items in both types of varieties and would thus concentrate in the upper intervals only.

Table 6.7: General nouns as R-1 collocates of *a/an* among top 500 words in all varieties

	BrE	AusE	NZE	AmE	SgE	IndE	PhilE	KenE
Total 'a/an + noun'	6585	5511	9801	3135	6840	5599	6160	5904
man	89 1.4%	58 1.1%	90 0.9%	53 1.7%	50 0.7%	50 0.9%	61 1%	119 2%
woman	52 0.8%	51 0.9%	78 0.8%	47 1.5%	33 0.5%	51 0.9%	62 1%	102 1.7%
child	51 0.8%	38 0.7%	37 0.4%	18 0.6%	56 0.8%	40 0.7%	37 0.6%	96 1.6%
person	39 0.6%	82 1.5%	59 0.6%	23 0.7%	65 1%	92 1.6%	82 1.3%	87 1.5%
girl	19 0.3%	23 0.4%	16 0.2%	10 0.3%	16 0.2%	23 0.4%	32 0.5%	21 0.4%
boy	0 0%	0 0%	0 0%	0 0%	10 0.1%	12 0.2%	19 0.3%	21 0.4%
object	12 0.2%	0 0%	13 0.1%	0 0%	0 0%	0 0%	26 0.4%	10 0.2%
idea	27 0.4%	30 0.5%	17 0.2%	15 0.5%	30 0.4%	0 0%	25 0.4%	11 0.2%
creature	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%
place	30 0.5%	45 0.8%	49 0.5%	24 0.8%	37 0.5%	26 0.5%	0 0%	0 0%
move	0 0%	0 0%	18 0.2%	0 0%	0 0%	0 0%	0 0%	0 0%
question	45 0.7%	0 0%	49 0.5%	26 0.8%	43 0.6%	43 0.8%	0 0%	60 1%

matter	64 1%	61 1.1%	79 0.8%	25 0.8%	48 0.7%	53 0.9%	93 1.5%	39 0.7%
thing	25 0.4%	33 0.6%	25 0.3%	10 0.3%	21 0.3%	27 0.5%	0 0%	23 0.4%
affair	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%
business	17 0.3%	0 0%	16 0.3%	14 0.4%	0 0%	21 0.4%	22 0.4%	0 0%
people	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	47 0.8%	0 0%
<hr/>								
total	470 7.1%	421 7.6%	546 5.6%	265 8.5%	409 6%	438 7.8%	506 8.2%	589 10%
<hr/>								

There are a few observable patterns in the percentages of the individual nouns. *Creature* and *affair* are not substantially used by all varieties. *Boy* is only substantially used by the Outer Circle varieties. *Girl*, on the other hand, is present in all varieties. It is possible that the more frequent use of *girl* is caused by its wider polysemy. As an example, *a girl* is used below to mean ‘a woman’. Were this a case of ‘a boy’ used as a substitute for ‘a man’, the meaning would not be equivalent.

144. Probably he 's a girl trapped in a man 's body <ICE-PHI:S1A-073#43>.

Move is not used substantially in all varieties except NZE. *People* is only substantially used in PhIE. There is a strong tendency in this variety to use *a people* as a reference to ‘a nation’. This is exemplified below:

145a. This is obvious to a government like ours that came to power by its means as

well as to *a people* like the Germans who attained complete freedom in the same way <ICE-PHI:S2B-027#84>.

- 145b. The translation, at the same time, aims to memorialize the resources of *a people* to heal its wounds, resources that are also portrayed in some of these texts <ICE-PHI:W2A-006#19>.

The patterns for the other nouns are vague, with no clear dividing lines between the Inner and Outer Circle varieties. Nevertheless when all the nouns are combined, the percentages show a slight tendency for the Outer Circle varieties to use more general nouns than the Inner Circle varieties. SgE defies this trend; with 6%, it is the second lowest user of general nouns after NZE. This finding recalls the use of general nouns with the definite article in which all Outer Circle varieties except PhilE showed a higher-use tendency. In the Inner Circle, AmE distances itself from the others by obtaining 8.5%, placing itself second after KenE overall. The lack of clarity in the overall trends observed in table 6.7 makes it difficult to say with certainty if the finding shows a parallel with the high amount of generality usually found in learner language. The fact that an Inner Circle variety (i.e. AmE) is also found to contain a high percentage of general nouns shows that native speakers, too, produce a large amount of generality in their discourse, although judging from its communicative efficacy, this use is apparently (communicatively) strategic and not aimed to compensate for more precise vocabulary.⁴⁵ Having said this, KenE is notable with its high percentage of 10%. With such an exceptionally large quantity of general nouns, is it justifiable to

⁴⁵ In addition, it has to be noted that AmE contains some dissimilar data in terms of sampling. If we discount this AmE finding, then it can be said with more conviction that the Outer Circle varieties tend to use more general nouns with *a/an*.

equate the amount of generality they produce with that associated with ESL learners? Such an argument remains difficult to uphold. To illustrate, compare the following uses of *a thing* in KenE and AmE:

146. The way I see it then is that it is obscene to turn <ea/>Harambee into *a thing* to raise money for personal gain. It is unforgivable to celebrate that you raised Sh100,000 for a wedding when you need only SH20,000. You need to have no conscience at all to turn the birth of a baby into a money-raising event <ICE-KEN:W2E018K>.
147. Yet the experience is the same: an ineffective drug exposing pregnant women to needless risk and sometimes even death. How is it that such *a thing* should happen <FROWN: B25 162-B25 164>.

Both examples are taken for the ‘editorial’ register. It seems that both writers have adeptly utilised the general noun to express the kind of meaning appropriate to their texts. In (164) it is used to characterise the *harambee* occasion (see Mahlberg, 2005: 108) while in (165), it packs all the information contained in *an ineffective drug exposing pregnant women to needless risk and sometimes even death* and creates the kind of cohesion identified by Halliday and Hasan (1976) in the next sentence. None of the uses appears so general as to end up as being vague. As regards specific nouns, there are less than 0.1% of them in each variety. As such they will not be discussed in detail.⁴⁶

⁴⁶ This also true for R-1 adjectives.

6.3.2 L-1 collocates of *a/an*

The figure below is a list showing the top 12 L-1 collocates of *a/an*. Only 12 are shown here because the words that are placed lower than the 12th position have fewer than ten occurrences and are thus not considered salient for our purposes.

Figure 6.9: Top 12 L-1 collocates of *a/an*

1	in
2	of
3	only
4	is
5	with
6	such
7	and
8	was
9	for
10	to
11	just
12	on

As in the case of the definite article, in the position first to the right of *a/an* we find an overwhelming number of prepositions i.e. *in*, *of*, *with*, *for*, *to* and *on*. The frequent use of NPs as a complement of PPs is evident although recall from the above discussion (see 6.2.3) that there are more definite than indefinite NPs that carry out this syntactic function.

Others on the list include singular *be*-verbs, showing their compatibility with the singular meaning of *a/an*. *Only* too reinforces this singular meaning while *such* is a frequent collocate as a result of the almost idiomatic use of *such a* as in:

148. And we cannot accept *such a principle* <ICE-PHI:S2A-067#32:1:A>.

The data show that there are no substantial differences in the L-1 collocates between

the varieties. On the whole we see little variation in the use of *a/an* at the two-word level.

6.3.3 Longer-span collocates of *a/an*

Figure 6.10 is a list of clusters shared by all varieties identified from the top 200 in each variety. Notice that there are many more overlapping clusters of *a/an* than that of *the* in figure 6.6 above. This may be an indication of a greater homogeneity among the varieties in terms of indefinite article use. Other than this, many of the words that are found to be collocates at the two-word level earlier are now observed to be operating as clusters, e.g. *a lot of*, *a little bit*, *as a result* and so on.

Figure 6.10 shows a few partitives which in turn function as non-numerical quantifiers. *A lot of*, *a number of*, *a group of* and *a piece of* appear to be common in all varieties. Of these, *a lot of* can take both singular and plural noun complements making it a very commonly used quantifier (Smith, 2009) and emerges as the most frequently used cluster in figure 6.10. Other ‘a + N + of’ patterns include those of *a kind of*, *a result of*, *a matter of* and *a series of*.

Figure 6.10: Clusters shared by all varieties among their top 200

a lot of	would be a
there is a	which is a
it was a	a matter of
to be a	more than a
a number of	we have a
there was a	a series of
this is a	is not a
to have a	a long time
it is a	there is an
a little bit	I had a
as a result	is a very
a kind of	he was a
and a half	I have a
a result of	that there is a
as a result of	a piece of
you have a	that is a
in a way	a group of
to be an	I was a
has been a	is a good

In chapter 2 (see 2.1.2) we saw that the indefinite article can be used non-referentially and play a descriptive/classifying role. This is the other dominant pattern of use found in the list. These include the ‘pronoun + *be*-verb + a’ pattern and others with a ‘*be*-verb + a’ stem. Some examples are *it was a*, *it is a*, *would be a*, *is not a*, *I was a*, *that is a*, *has been a* and so on. *There*-sentences also typically require the indefinite article. The verb *have* which is known to usually take an NP complement headed by a determiner (in this case, *a/an*) is shown to collocate with the pronouns *you*, *we* and *I*, showing the central role that these pronouns play in an interlocutor’s language. *In a way* is a PP that has become a very common adverbial with *way* often being used metaphorically (see 6.1.1). Finally the above table shows that the broadly ‘positive’ adjective *good* is also a frequent collocate of the indefinite article.

Table 6.8 presents the clusters of the indefinite article according to the same four categories used for *the* in table 6.6 above.

Table 6.8: Clusters of *a/an* in Inner and Outer Circle varieties

In two or more Inner Circle varieties only	In two or more Outer Circle varieties only	In one Inner Circle variety only	In one Outer Circle variety only
<p>BrE, AusE and NZE I've got a a range of a bit more we've got a he's got a and it's a was a bit be a bit</p> <p>AusE and NZE have a look at for a couple of hell of a</p>	<p>none</p>	<p>BrE only a single currency going to have a a much more to provide a well it's a so it's a have had a in a minute I mean a in fact a a certain amount it's a sort I enclose a may be a</p> <p>AusE only they've got a she had a is that a hour and a an increase in a cup of just a little a sort of a it's a good have a bit days a week but it's a kilometres an hour and a lot a hundred percent have a bit of she's got a and a lot of a whole lot to have a look it's been a and as a got a lot there has been a got a bit a few weeks</p>	<p>SgE only going to be a you need a you are a can be a one and a one and a half have a lot of a by election take a look at are a lot example of a two and a half to buy a is there a think it's a quite a number quite a lot of much of a ...</p> <p>IndE a member of is in a a function of is a kind was in a I am a she is a is a great to form a India is a as a teacher as a matter of only a few in a particular a man who a time when it is a very has got a such a way have a very was not a</p>

		<p>had a bit it's a bit of had a bit of</p> <p>NZE only a New Zealand a hell of a wee bit may be a was a good has had a and there was a a really good had a good and there's a which has a a new Zealander</p> <p>AmE gonna be a only a few a handful of a picture of of a new all of a sudden a pair of a total of a way to all of a he is a a bunch of up to a ...</p>	<p>...</p> <p>PhiE to become a are a lot is there a a case of there are a lot are a lot of can be a a few minutes an average of to have an have a lot of as a people be a very a strong republic ...</p> <p>KenE a point of a lingua franca an assistant minister who is a a view to an hon member ...</p>
--	--	--	--

6.3.3.1 Clusters of *a/an* in the Inner Circle varieties

In the Inner Circle we find that the verb *has/have* tends to be contracted more often when preceding the indefinite article in BrE, AusE and NZE. We also find that the word *bit*, which usually functions as a quantitative partitive, is used more frequently in these varieties.

Concordance 6.9 shows all the 17 lines of *I've got a* found in the Outer Circle varieties and 20 selected lines of *I've got a* from the Outer Circle varieties. It can be observed

that speakers of the Inner Circle varieties use the phrase in a wider variety of ways. The indefinite article in *I've got a* is often part of a quantifier, either numerical or non-numerical, in these varieties. We see that this is the case in the Outer Circle too, e.g. *I've got a few, I've got a couple*. However, the difference is that in the Outer Circle such uses are infrequent. *Friend* seems to be a frequent object of the verb *got* in both groups of varieties. Nevertheless the concordance shows that Inner Circle speakers also use *I've got a* to indicate possession of a larger variety of entities.

Concordance 6.9: 17 lines of *I've got a* in Outer Circle varieties and 20 selected lines of *I've got a* from Inner Circle varieties

Outer Circle varieties

Also tomorrow	I've got a specially arranged
really getting serious now!)	I've got a few friends up at T
mind having her address) but	I've got a friend living in th
oice please excuse me because	I've got a very bad cold at t
was and uh I was hoping that	I've got a place there
say this because uhm I have a	I've got a very big title for
Well	I've got a feeling that Soe na
Not too good for me	I've got a sore finger uh
and there's no point saying	I've got a legal mortgage it i
Well	I've got a bunch of people who
	I've got a friend uh I don't k
	I've got a couple of newspaper
forward and asked me whether	I've got a change for a dolla
No	I've got a different
Ya but the	I've got a small problem here
So Gorbachev said to Bush:	I've got a very simple process
Tomorrow	I've got a tutorial to give so

Inner Circle varieties

and	I've got a couple of queries o
universe S2A-055(noone):4	I've got a couple of S2A-05
e dictionary project, so yeah	I've got a few days down here
Right well	I've got a few here but Greeds
erfect S1B-020(A):261 I've	I've got a few of them here
surname was um Nguyen 'cause	I've got a friend in Sydney ca
l drop it back for paulie cos	i've got a friend of mine that
(continued)	
yeah cos	i've got a friend um tut
In my In my cupboard	I've got a lot of rejections,
e	I've got a lot of students peo
sent Abstract and abstract Um	I've got a lot of them S1B-

(continued)

	Since I've got a match on Sunday at
to sleep on the spare	i've got a mattress thing i ma
Not that	I've got a miniskirt you under
enormous S1A-033(C):128 So	I've got a whole chapter on ho
	, I've got a whole load of my ow
If you give me a tape	I've got a tape to tape and I
one):59 dotted line She said:	I've got a terrible headache a
generation here Uhm you know	I've got a son that, well I'm
d guess what S1A-096(B):93	I've got a sore stomach too so

Concordance 6.10 shows all the ten lines of *we've got a* and all the seven lines of *he's got a* in the Outer Circle varieties. It also shows ten selected lines of *we've got a* and ten selected lines of *he's got a* from the Inner Circle varieties. Unlike what we saw for *I've got a*, these two phrases have a different motivating factor for the observed frequencies in the two types of varieties. One of the characteristics of the NP complement of *have got a*, which can't be shown in the limited contexts of the concordance lines, is that it tends to receive further description which in turn takes the form of any of the following: premodification, postmodification or elaboration in coordinated or uncoordinated sentences. Below is an example of this description being a postmodification:

149. It's gonna be unreal And then *we've got a video that Firstname1 made as well* a Student Life one <ICE-AUS: S1A-013(B):298>.

Examining the data in concordance 6.10, one detects a slight tendency in the Outer Circle varieties to use postmodification as opposed to premodification to provide the required description to the NP. This is illustrated by the first four lines of *we've got a* and the first three lines of *I've got a* in the Outer Circle group. In contrast, in the Inner Circle there is a good mix between premodification and postmodification for the NPs used. It is possible that the frequency of the use of the indefinite article via these

phrases increases in the Inner Circle group because of their greater flexibility in this regard.

Concordance 6.10: 17 lines of *we've got a* and *he's got a* in Inner Circle varieties
and 20 selected lines of *we've got a* and *he's got a* in Outer
Circle varieties

Outer Circle varieties

e PAP is now in government so we've got a mandate to govern
You would realise that we've got a printer that is of
ere is no fight because here we've got a interest that is u
About a quarter to five We've got a tape we would want
w We've got a bit more hands-on
Uh we've got a new product strate
We've got a declaration of hum
Now we've got a settling unit and
arning environments sometimes we've got a way of constructin
by virtue of the fact that we we've got a centre for nuclear
's graduated in Marketing and he's got a job in Australia wh
him will try to get over him he's got a few players waiting
He's got a new one and he's qu
But he's got a hand in everything
I think he's got a point The point is
He's got a David Brown doesn't
l be done for you like I know he's got a very good son It's

Inner Circle varieties

now then dave we've got a bloody spark
Well I think perhaps we've got a different position
hey can hear it so um yeah so we've got a broadening of indu
Well we've got a great range now th
e're not yet out of the woods we've got a long way to
rofit share S1B-039(B):256 We've got a whole range of oth
o to the shop to buy milk and we've got a great variety to c
of the breed S2A-012(A):65 We've got a female that's oozi
on tuesday we've got a function of say th
But we are students so we've got a right to say ourse
tes and he comes back out and he's got a big smile on his fa
at the same time he's got a huge profile is he
he's got two uh two bowls and he's got a straight hit at it
lowing it whereas the freshie he's got a very long narrow sn
etimes in these type of races he's got a withering kick, but
Except for he's got a bit of a tash now,
I mean he's got he's got a lot of problems at
as telling me last night that he's got a mate in the
d down a budget like that and he's got a petition going to
moore showed why he's got a reputation for havi

Concordance 6.11 shows all the 24 lines of *a range of* found in the Outer Circle group and 25 selected lines from the Inner Circle group. Inspection of the concordance reveals that in the Inner Circle there is a possibility that the frequency of the *a range of* phrase is influenced by the greater variety of prepositions used by their speakers. Whereas the Outer Circle varieties' speakers are shown to use the phrase as the object of PPs headed by *for, of, with* and *within*, Inner Circle varieties' speakers additionally use *by, from, on, over, across* and *through*. Also, while in both groups there is a variety of verbs found to collocate with the phrase, there is a wider range found in the Inner Circle.

Concordance 6.11: 24 lines of *a range of* in Outer Circle varieties and 25 selected lines of *a range of* from Inner Circle varieties

Outer Circle varieties

<p>just show we just give him a rry we can we can accommodate now big enough to accommodate million hectares, as well as r sub-contractors as well as ge adding that he has brought ge, adding he has also bought So you can consider why children have to develop The Cashcards can be used for conies it is possible to grow ions system manned by NAA has We are giving them in i motors coming in ldbirth; (d) availability of The Ateneo provides arbitrary meaning but rather stingly techno as a prefix to There was ection-finding equipment with f non-radiative channels with ustomers to make calls within It has stabilized within</p>	<p>s. A range of dyes with different a range of tapes from the Dra a range of opinions and I hav a range of views In our early a range of demand management s a range of uh consultants that a range of guitar effects for a range of guitar effects for a range of variables related t a range of activities. a range of purposes. a range of shade orchids. a range of nearly 1,000 miles a range of only ten percent to a range of ratings and designe a range of choice appropriate a range of other services to s a range of acceptable or appr a range of terms and activiti a range of mountains in front a range of 100 - 150 miles to a range of lifetimes. a range of a public based stat a range of thirty-seven to thi</p>
---	---

(continued)

Inner Circle varieties

mediation or arbitration by a range of people, such as:
budgets er was accompanied by a range of social polic
for system interruption from a range of causes, for example
will maintain equipment from a range of hardware vendors at
T he times are typical of a range of kernels that were i
18 The weights and volumes of a range of pebble sizes were a
n of different antagonists on a range of agonists.
soon proved themselves on a range of other species.
phore, which can respond over a range of light intensities c
points remained constant over a range of temperatures,
be sufficient progress across a range of genuine human right
about force statistics across a range of variables (details
rs into rural schools through a range of incentives. W1A-
41 Each model was put through a range of technical and compa
gn fault which had existed in a range of instruments for 10
201 It can also be ordered in a range of larger calibres suc
Packard 2D laser printer with a range of scalable fonts und
atrons constantly amused with a range of special-theme oc
employees at a point within a range of pay rates, based up
copy of solid materials using a range of nuclei, and magnet
and Setchell (1974) recorded a range of hopping frequencies
remainder of this chapter is a range of successful baits an
tem or society which includes a range of values from
al Types which would describe a range of ways in which the B
r between the environment and a range of alternative develop

In the case of *and it's a*, it is not easy to identify the cause(s) that lead(s) to the higher frequency of this phrase in the Inner Circle varieties. Concordance 6.12 shows all the 14 lines of *and it's a* found in the Outer Circle group and 14 selected lines of the same cluster from the Inner Circle group. Examining the concordance, one can only say that Inner Circle varieties' speakers have a greater predilection to use this phrase than their Outer Circle counterparts. Apart from this, one can detect a slight tendency by the Inner Circle speakers to use two nouns, *matter* and *question*, as the nominals in the resulting NPs. *Very* appears to be the only adverb that is substantially used to modify the adjectives used in the nominals. As mentioned earlier, other preferences by speakers of the Inner Circle varieties are difficult to ascertain.

Concordance 6.12: 14 lines of *and it's a* in Outer Circle varieties and 14 selected lines of *and it's a* from Inner Circle varieties

Outer Circle varieties

s I take over his empire yeah And it's a business empire for
 A second charge comes along and it's a fixed charge Now th
 r of uh Evans only for awhile and it's a goal kick to Saraw
 And it's a it's I see it becau
 hy is this thing so expensive and it's a life-time drug Why
 it's it's thirteen Hong Kong and it's a lot of rice very l
 re is business in Parliament, and it's a major attraction.
 she lunges towards the walland it's a new record time by
 nother goal kick to Singapore and it's a question of time as
 coming from home Okay And it's a thriller Well
 he miss kick by Chai Siew Yai and it's a throw to for Singa
 And it's a very valid question
 Yes Kenyans for treatment and it's a very expensive cent
 etermines the state of health And it's a very tricky uh rela

Inner Circle varieties

rd and it's a matter of law,
 er and it's a matter of debate in
 tinction is very very blurred and it's a question as to what
 got a tiger by the tail there and it's a question of
 if you cut across a fillet and it's a very economical cut
 he Kung S2B-041(noone):212 And it's a very sharp contrast
 her the offence was committed and it's a reasonable doubt ha
 5(A):88 Yeah you call them up and it's a recorded message
 and having a look and it's a pain in the neck
 e company goes down the drain and it's a partnership each pa
 her it be night or day and ah and it's a great disappointmen
 035(B):142 Mind you so is sex and it's a hell of a lot warme
 ke everyone's thesis is there and it's a blue cover with gol
 one with profit maximisation and it's a choice of choosing

The situation is slightly similar for *be/was a bit* where the reason(s) underlying the Inner Circle speakers' choice for using these two clusters is/are largely unclear. However, there is a perceptible semantic/pragmatic difference between the two groups of varieties. The *be a bit* and *was a bit* clusters are used mostly with negative-meaning adjectives and adverbs or sometimes with positive-sounding adjectives or adverbs but rendered by the speaker as disapproving or critical of the situation. In addition, when occurring in an imperative, the use of these phrases serves to foreground the positive

value of the modifiers with the effect of turning the meanings of the adjectives (or sometimes adverbs) into something of a necessity. Upon closer inspection, one can detect a slightly greater tendency of the Inner Circle varieties' speakers to go against this norm by using positive adjectives with the positive meaning unchanged throughout the utterance. This is illustrated by the first five lines and the last four lines in the Inner Circle section of concordance 6.13. On the contrary, such a pattern of use occurs only once in the Outer Circle data, with the use of *was a bit* but not with *be a bit*. The last line of the Outer Circle section demonstrates this pattern of use.

Concordance 6.13: 47 lines of *be/was a bit* in Outer Circle varieties and 50 selected lines of *be/was a bit* in Inner Circle varieties

Outer Circle varieties

But also be a bit adventurous
you can read that uh it might be a bit blurred there for the
on lights I think you have to be a bit careful with them bec
rd uh it proved to be a bit cough troubl
terms of getting water it'll be a bit difficult alright
son and Johnson and uh might be a bit difficult to get in u
vinces uh then I think it may be a bit difficult
e on yes who'll do that Let's be a bit gender sensitive now
So the level of demand would be a bit higher
well surely mean that they'll be a bit more cautious with t
element of really seeking to be a bit more scientific about
move commercial but it should be a bit more balanced than w
we're thinking the dates may be a bit of a problem
iwali is supposed to be a bit of gamble a bit of
d return uh can prove to be a bit of disadvantage in i
So there has to be a bit of a balance
is that there was there will be a bit of slowdown but that
at not to say: Yes, I used to be a bit sensitive in those da
beautiful tree goddess might be a bit spoilt because part
But I feel that it will be a bit too late so I'm
o have difficulties do try to be a bit uh adventurous a bit
amme I know you are trying to be a bit uh very cautious in t
hey get married and uh it may be a bit uh too much of a hig
ually I want to mention but I was a bit afraid that I'll be
of cover note and there I was a bit busy with
I was a bit confused with the yo
a relevant objection Njeri I was a bit disappointed myself
ned the manuscript I think it was a bit discouraging without

(continued)

of that system so at first I was a bit disoriented
Guess I was a bit enthusiastic
I 've heard that he was a bit intimidated by me
He was a bit like Manfred actually
t a throw out of that one who was a bit of misunderstanding
Now there was a bit of uneasy concern on
There was a bit of a push on Zubir I
dialogue session at Moulmein was a bit of a threat when I
There was a bit of drama on the day
I was I was a bit slow lah
s that occurred in La Salle I was a bit surprised because I
fe played by Anne Karuiki who was a bit too mild in the trou
That ball was a bit too soft by Melanie
I mean it was a bit uh a bit uh
arlier whole day he was a bit uncomfortable
h and but you know it was a bit unexpected because I
Sarita was a bit unsteady there in an
my goodness gracious me that was a bit word
Paolo was a bit athletic himself, al

Inner Circle varieties

ested in and so that tends to be a bit easier and ah they kn
so that'll be a bit exciting
rts school then she'll be a bit freer to to
you are inspected expected to be a bit funny and I as I was
l-de-sac S1A-005(B):5 He'd be a bit younger than Jeff wou
t she meant and that they may be a bit alienated by this so
yeah perhaps she'd be a bit better looking if she
agine the first few days will be a bit hectic, during which
he loam content ' content can be a bit iffy as a result.
B-080(A):313 Um it's going to be a bit more expensive to do
t i think sherry might be a bit much for a picnic so
Would be a bit of a bad deal wouldn'
all right well a is gonna be a bit of a problem
but it sounded like it might be a bit too late for u
to her that half shares might be a bit unfair when I was the
George, who is BMW Fairbanks was a bit cool at first.
Fran was a bit depressed when she h
07(A):12 Um I don't know, he was a bit far away for me to s
you know i was a bit hesitant at first bu
ckers and the water pressure was a bit high too, which left
Besides, it was a bit lonely, a bit embarr
that one S2A-004(A):66 It was a bit long so Martin holds
e a bit better looking if she was a bit more lively i
i suppose it was a bit much taking four kid
so I was a bit of a coward,
and who's firstname2's father was a bit of a bastard really
and then there was a bit of banging and
it was a bit of a cheat
The lady who ran the place was a bit pissed off because w
e, and come back up again. He was a bit puzzled about this s
was tempted to say that but i was a bit reluctant to
um i just thought it was a bit ridiculous all the p

(continued)

yeah but he was a bit round the twist was n
I was drinking a Coke and I was a bit shaky and a bit a bi
yeah actually that was a bit silly of me
S1A-092(B):290 One of them was a bit slow with their meal
S1A-007(B):256 Even though it was a bit sort of weird being
Uh I think that was a bit unfair,
I was a bit unwilling to, make t
he started singing his voice was a bit weak S1A-075(A):4
and it was a bit wobbly because he wa
e finish because er he was a bit worried about the fa
sanyo 's instruction booklet was a bit confusing in places,
my father was a bit conservative in that
the Prime Minister, Mr Hawke, was a bit cool on the idea.
ich case i sort of thought it was a bit awful to
en S2A-010(A):130 Well I I was a bit of a fan of Kyote ac
that was a bit of a laugh e
ool but because I screamed it was a bit better S1A-019(B)
i admit there was a um there was a bit of unity that

The last cluster that is more substantially used by the Inner Circle varieties is *a bit more*. Concordance 6.14 presents all the 22 lines of this phrase in the Outer Circle varieties and 25 selected lines from the Outer Circle varieties. The concordance shows that both groups of varieties use this phrase mainly as an adverbial to modify adjectives. There are also instances of it being used to modify adverbs as well as PPs. The difference between the two types of varieties is that the Inner Circle speakers' use of this phrase in each of the roles specified above is more wide-ranging. We can see this especially in its use with prepositions. In the Outer Circle varieties, the only prepositions used are *about* and *for*. In their Inner Circle counterparts, we find *about*, *round*, *under*, *up*, *into* and *of*. As a modifier of verbs, its use is more prolific in the Inner Circle. This is evidenced by the variety of verbs found in this group. While in the Outer Circle the use of *be*-verbs overwhelms the other kinds of verbs found there, in the Inner Circle *a bit more* is more frequently found to modify lexical verbs such as *hear*, *do*, *have*, *add*, *play*, *travel*, *climb*, *show*, *talk (about)* and so on. Finally the conjunction *than* is also used by speakers of the Inner Circle varieties to emphasise the

comparative adverbial meaning of *a bit more*. This pattern of use is absent in the Outer Circle varieties.

Concordance 6.14: 22 lines of *a bit more* in Outer Circle varieties and 25 selected lines of *a bit more* in Inner Circle varieties

Outer Circle varieties

Trying to read a bit more
 is is where I have taught you a bit more about thinking and
 y to be a bit uh adventurous a bit more adaptable and be pr
 e commercial but it should be a bit more balanced than what
 the derby on Wednesday night a bit more bearable.
 l surely mean that they'll be a bit more cautious with this
 ly arrange at a venue that is a bit more conducive rather t
 o stress and others which are a bit more diffused ah
 stories, though - - they took a bit more effort.
 ough sometimes you you you see a bit more for maybe politica
 We've got a bit more hands-on later on
 It's more a bit more judgemental So he a
 You seem to be getting a bit more mileage out of that
 I think I'd prefer something a bit more modern," Wen Ei sa
 fast because you kind of have a bit more road-sense already
 ement of really seeking to be a bit more scientific about t
 ens they can start uhm being a bit more serious about the t
 Then you earn a bit more something like that
 t as time passed, it had been a bit more spread out and in
 A bit more transgenic now
 so much oyster sauce you use a bit more uhm all sorts of t
 ribute these two masterpieces a bit more widely. The writt

Inner Circle varieties

i actually want to hear a bit more about the role play
 opmental theory we might hear a bit more about the child see
 arse sieve i should really do a bit more but we're ru
 so I thought I'd better have a bit more but I've gotta fill
 around as well then that adds a bit more cost as well S1B
 we've added a bit more detail
 easant messages across a bit more kindly
 listeners who i'll talk about a bit more later as well to se
 hey'll come back now and play a bit more round home
 I'll just turn it around here a bit more S1A-001(A):53 Th
 unge room S1B-080(A):48 Um a bit more storage by the side
 t of view they do have a bit more strength on the ben
 Well it's a bit more than a Land Rover
 a film will spell things out a bit more than a play because
 1A-025(B):275 She's travelled a bit more than Kylie has S
 now with probably a bit more time and effort whi
 dren have got that white line a bit more under control
 hy S2A-055(noone):148 It's a bit more uniform across the
 030(H):80 And I got to paddle a bit more up the front all th

(continued)

i thought oh i'll just climb a bit more you know and i'll b
urings and herbs if you fancy a bit more of a culinary cook'
they're a bit more into boys
he end of the week um er with a bit more rain and stuff
we can show a bit more guts in our dealing
p getting nothing if he isn't a bit more forthcoming!

6.3.3.2 Clusters of *a/an* in the antipodean varieties

Table 6.8 shows that several clusters are used substantially in AusE and NZE only.

These are *have a look at*, *for a couple* and *hell of a*. It is possible that the preference for these phrases has influenced the use of *a* in these two varieties.

Concordance 6.15 shows all the 37 lines of *have a look at* found in the Outer Circle varieties and ICE-GB and 40 selected lines of this phrase in ICE-AUS and ICE-NZ. The concordance shows that the phrase is frequently used as an imperative, which also includes the use of the first person inclusive *let*-imperative (see Collins, 2004). *Have a look at* also frequently collocates with *to* and together they form an infinitival clause. There is no clear difference between the antipodean varieties and the others except for the fact that speakers of the former varieties obviously prefer the construction with the light verb *have*. It is possible that speakers of other varieties simply use 'look' or another light verb construction 'take a look at' for the same meaning.

Concordance 6.15: 37 lines of *have a look at* in the Outer Circle varieties and ICE-GB and 40 selected lines of *have a look at* from ICE-AUS and ICE-NZ

Outer Circle varieties and ICE-GB

Have a look at some of these a
 Have a look at that one again
 Have a look at the next pair o
 and have a look at my Dayspring al
 tandard students just go and have a look at them
 So if you want you can have a look at the book
 And then he can have a look at them and see wh
 For example have a look at this table
 Can I have a look at it
 Uhm I'll have a look at that before I g
 s had recommended, uh to just have a look at just so that,
 a look at that one, and let's have a look at address,
 Let's have a look at some of these w
 Let's have a look at that date
 Well let's have a let's have a look at that one, and l
 So let's have a look at the latest movi
 Let's have a look at the list,
 Let's have a look at the next pair o
 Let's have a look at the next two se
 Let's have a look at the next pair o
 Let's have a look at the next two se
 Let's have a look at the next word
 Right now let's have a look at this particular
 Let's have a look at your throat jus
 Now have a look at this bedsheet
 But we quickly have a look at the meaning
 rs can I ask your Lordship to have a look at page forty-five
 nnection it is instructive to have a look at the figures fo
 ing parties of kids around to have a look at the site,
 tten out of the room so as to have a look at the vehicle whi
 I meant to ask them to have a look at them the chits
 's a I mean you're welcome to have a look at this one but I
 have people who go around to have a look at what is happeni
 Did you uhm uh have a look at her when she ca
 You sit there and you have a look at it
 Can you have a look at page twenty
 Could you have a look at the damage repo

ICE-AUS and ICE-NZ

have a look at him here he
 ou like S2A-053(noone):193 Have a look at that S2A-053
 i'll have to go up and have a look at the plai
 said to me kevin come and have a look at the the paint o
 f the Penrith store there and have a look at the way they're
 to take them away and have a look at them
 find celebrating the day and have a look at them in there
 ids getting right into it and have a look at this S2A-010
 ou if you um go back home and have a look at your bibles if

(continued)

t to have a look you'd better have a look at it S1B-070(C
other little brochure you can have a look at if you're inter
the table and you can have a look at them later
let's just have a look at that cos the ot
S2A-024(A):29 You might just have a look at that S2A-024
ers S2B-039(A):21 And just have a look at the flames that
o if anybody doubts that just have a look at this S1B-044
S2A-002(B):63 And now let's have a look at our own Kylie S
okay let's have a look at part c
now let's have a look at serving some wh
S2A-032(noone):172 Now let's have a look at some of the par
way towards the mat and let's have a look at that gam
nd while that's cooking let's have a look at the house S2
can change the view and then have a look at something like
):204 The employers will then have a look at them and select
hem he was more than happy to have a look at
to have a look at
ng i was wanting to do was to have a look at a contin
6(A):20 I ask Your Worship to have a look at ah the defendan
i went to have a look at her wedding cak
taubmans people in to have a look at it
do you want to have a look at it 12:40
is is planning to have a look at people's
if we have a look at the on page nin
r ah your scores and we we'll have a look at what the overall
yes but then if you have a look at a few lines dow
S2A-060(A):109 And if you have a look at Effie now the f
S1B-067(A):21 Alright Did you have a look at his face after
night in Valdez ('Val-Dees'), have a look at the southern e
07 And speaking of energisers have a look at that S2A-010
S2A-018(B):207 We're gonna have a look at it now actually

The situation is similar in the case of *for a couple of*. Concordance 6.16 presents all the lines of this phrase in ICE-AUS and ICE-NZ, ICE-GB and the Outer Circle group. ICE-GB and the Outer Circle group are separated this time to show clearly how a pattern of use is present in all three groupings/types of varieties (i.e. antipodean, Outer Circle and BrE) but the tendency for it to occur is greater in AusE and NZE (as we have observed for *have a look at* above).

Concordance 6.16: 14 lines of *for a couple of* in the Outer Circle varieties, 9 lines of *for a couple of* in ICE-GB and 39 lines of *for a couple of* in ICE-AUS and ICE-NZ

Outer Circle varieties

She won't be back for a couple of days.
 se , Mum has been sick for a couple of days but now s
 If it persists for a couple of days then you
 ith a bed for a hundred pesos for a couple of hours.
 He let the gray loose for a couple of minutes: it go
 I was a branch manager there for a couple of months before
 ke to have the English nurse for a couple of months"
 for a couple of more days
 I heard she 's been there for a couple of times
 Uhm and they actually ask for a couple of uh people a co
 e in the States for a couple of weeks
 But after deferring for a couple of years they gav
 a certain lady We've gone out for a couple of years by the w
 ey 've been in the Final Four for a couple of years now and

ICE-GB

rough the rest of the war and for a couple of years after it
 parents lived in Bournemouth for a couple of years
 he paper, to acting as editor for a couple of weeks when Bos
 that she could live frugally for a couple of days on her ea
 ged to meet Mary this morning for a couple of hours before s
 They buy the books say for a couple of pounds, and th
 we would stop for a couple of minutes to pla
 ropped back now having led uh for a couple of laps,
 es two or three times a week, for a couple of hours at a tim

ICE-AUS and ICE-NZ

iver's head, with ample space for a couple of communications
 what's that one overseas trip for a couple of ministers
 just sort of being there for a couple of days
 (A):102 We can go without him for a couple of days S1A-08
 S1A-065(A):83 But that's only for a couple of days S1A-06
 to fly to Canberra with them for a couple of days - the las
 S1B-080(A):35 Um get room for a couple of filing cabinet
 't get much of an idea for a couple of furlongs as th
 hey let us stay after closing for a couple of games
 the catamaran going out there for a couple of hours and then
 hat and that's where you stay for a couple of hours and they
 old him I said we'd only play for a couple of hours anyway
 now that should marinate for a couple of hours but er a
 aid he'd be only able to play for a couple of hours S1A-0
 you can get guitars you know for a couple of hundred dollar
 whakatane when i left school for a couple of months
 o overseas next year and stay for a couple of months S1A-
 (B):229 Some new guy's coming for a couple of months S1A-
 she said you need to leave it for a couple of months until
 ess so I found myself looking for a couple of months work.

(continued)

I think we need to be looking for a couple of new board memb
He usually just stays away for a couple of nights.'
uess i'd been in the lectures for a couple of of times but
john bracewell try him for a couple of overs pull him
ith picking up a \$900,000 tab for a couple of paintings tha
tes. "It didn't look too good for a couple of seconds. W2
sorry interference on tape for a couple of seconds
nd in Britain who'd been done for a couple of skinny joints
now i've chosen these people for a couple <.>of tape
who got a cheque in the mail for a couple of thousand dolla
actually i haven't seen him for a couple of we
ou know y if you go up for a couple of weekends go fo
98(A):83 They'll only be away for a couple of weeks S1A-0
to spin out the cassette tape for a couple of weeks until ou
do we have to go to australia for a couple of years
done half done inhales for a couple of years n
guy's been living in Majorca for a couple of years you know
e've only been looking for it for a couple of years."
yway we'll start in australia for a couple of years

According to the above concordance, (*for*) *a couple of* is mainly used in all varieties to quantify periods of time e.g. days, weeks, months, years, seconds, etc. To a much lesser extent in BrE, but not the Outer Circle varieties, it also used to quantify amounts of money and certain other nouns. As for AusE and NZE we see that all these uses are available but in regard to quantifying money and other nouns, it is used more frequently than in BrE.⁴⁷ Speakers of AusE and NZE are found to use the quantifier quite substantially with nouns other than those representing time and money, e.g. communication sets (from the first line of the ICE-AUS and ICE-NZ section), ministers, filing cabinets, members, paintings, overs (cricket matches), etc.

The last cluster found to be substantially used in AusE and NZE is *hell of a*. Concordance 6.17 shows all the 7 lines of the phrase in the Outer Circle varieties, all the 7 lines of the phrase in ICE-GB and all the 24 lines of the phrase in ICE-AUS and

⁴⁷ Smith's study (2009: 177) shows that the tokens for *a couple of* (but not *for a couple of*) are found in ICE-AUS, ICE-NZ and ICE-GB in the following ratio: 223:199:126. These figures are consistent with the observation being made here.

ICE-NZ. This quantifier-looking phrase is usually used informally as an adverbial. It has a role of an intensifier. It is alternatively spelled as 'helluva' (though this is not found in the data).

Concordance 6.17: 7 lines of *hell of a* in Outer Circle varieties, 7 lines of *hell of a* in ICE-GB and 24 lines of *hell of a* in ICE-AUS and ICE-NZ

Outer Circle varieties

lot of common sense involves hell of a lot of number crunch
But it involves hell of a lot of common sense
company which actually invests hell of a lot on their employe
It is clear that he is in a hell of a hurry to do somethin
es; that I could give you one hell of a bear hug!! May be yo
who, Inez believed, was one hell of a masochist for sticki
use it 's one of uh it 's one hell of a job to be married an

ICE-GB

s sense, and you gave him one hell of a beating.
it's it's going to save us a hell of a lot of time because
have an intention of doing a hell of a lot more from now on
he moment but it's taken me a hell of a long time,
It's a hell of a lot faster than anyt
nty-five pounds but there's a hell of a lot of information T
He's taking, a hell of a long time

ICE-AUS and ICE-NZ

mean he doesn't really make a hell of a good job of it
and got about 50 feet up - a hell of a height for an 11-yea
1b-073(A):143 Um it is been a hell of a job doing it let me
you know a hell of a long time
now jenny never used to say a hell of a lot about him
but there won't be a hell of a lot around of course
it would have taken a hell of a lot of work
afternoon when there isn't a hell of a lot to do and ah
5 Like S1A-042(B):126 Oh a hell of a lot S1A-042(A):12
009(A):170 You have to sell a hell of a lot of books to reac
8(noone):75 He went through a hell of a lot of ammunition. H
Mind you so is sex and it's a hell of a lot warmer S2B-03
but science is in fact a hell of a lot more creative be
ven though it's gonna cause a hell of a lot of grief to
ast and there aren't a hell of a lot of young scienti
flounder and there must be a hell of a lot of them from the
going to read about a hell of a lot more and hopeful
g it would make life a hell of a lot easier
t the supermarket there was a hell of a lot of people
oh no not a hell of a lot um
you don't know much I'm in a hell of a mood today S2A-05

(continued)

playing cards having a hell of a nice time
off down there and caused one hell of a scene.'
we've had a hell of a time over the last h

One of the things that can be immediately seen in the above concordance is that the Outer Circle group contains tokens of *hell of a* without being prefaced by *a*. Upon inspection, these four lines are found to originate from ICE-SIN. This appears to be the only case in this chapter in which a cluster leads to the omission of an article. Cases like this will be discussed further in the next chapter.

Most tokens of *hell of a* in ICE-GB are found to collocate with *lot*. This is reflected in greater proportion by ICE-AUS and ICE-NZ (but not the Outer Circle group). Sometimes the phrase is found to act like an adjective by directly modifying nouns e.g. *masochist, job, beating* (cf. 'helluva'). Cases such as these are found to be more frequent in ICE-AUS and ICE-NZ. In the Outer Circle, there is a tendency to replace the indefinite article determining *hell* with *one*. This pattern is found once in ICE-AUS and ICE-NZ. Therefore we can observe that all the patterns of use identified above (except for omission of *a* in ICE-SIN) are available in each group of varieties or variety. However, it is clear that the antipodean varieties have a predilection for this phrase as evidenced by their higher frequency both in the number of overall tokens and the number of tokens for each use identified.

6.3.3.3 Clusters of *a/an* in individual varieties

Table 6.8 shows that the varieties have a large number of clusters that are substantially used in each of them. Some of these clusters, however, are found to occur almost

equally frequently (with a difference of one or two tokens; one reaching or exceeding the cut-off point of ten and the other with only one or two tokens below it), e.g. *a hell of* in AusE and NZE. To ensure that only the most salient clusters are chosen to represent the individual varieties, the analysis will be limited to the top two most frequent clusters in each of them. The patterns that are representative of their variation are shown in table 6.9.

Table 6.9: Clusters of *a/an* in each individual variety vs. others

Cluster	Cluster pattern(s) in own variety	Cluster pattern(s) in other varieties
<p>BrE a single currency</p> <p>going to have a</p>	<p>hat could in time become a single currency urrency could evolve into a single currency but onl ng a decision on joining a single currency until a d to dismiss the idea of a single currency ity's transition towards a single currency. (24 tokens)</p> <p>Well they're going to have a hard time work Well we're going to have a fa family do a Aren't you going to have a drink of anyth You're going to have a good time here (21 tokens)</p>	<p>ign policy and introduce a single currency by 1999 (1 token)</p> <p>we're going to have a scrum Christmas, our church is going to have a service d it 's uh I 'm sure I 'm going to have a very fun Ya wow wee she's going to have a lot of br (42 tokens)</p>
<p>AusE they've got a (27)</p> <p>she had a</p>	<p>C):117 Yeah 'cos and now they've got a double stor s to get around to it so they've got a fair bit of -014(A):33 Yeah well see they've got a gang of the t Budget wholesale meats they've got a great quail (27 tokens)</p> <p>ectures so she she had a a few blunt wor Yes er she she had a little bit of a knee twist -041(C):94 So she had a practised skill -007(noone):31 She had a rude shock comi (28 tokens)</p>	<p>They've got a thing which they've got a software pa let the actors down when they've got a song or a d I honestly don't think they've got a problem (31 tokens)</p> <p>Because she had a bad experience She had a lot of swelling Caroline had told him she had a daughter with p hat is very interesting. She had a birthday party (60 tokens)</p>

<p>NZE a New Zealand</p> <p>a wee bit</p>	<p>very thrilled that it's a new zealand company and birth, being married to a New Zealand citizen bef whi which is a new zealand certificate two points a record for a new zealand team 46 tokens)</p> <p>nansen bottles are a wee bit complicated ckled and of course just a wee bit of sloppy play t seems to be pushing it a wee bit to ask what eth it got me going a wee bit yeah but i s (29 tokens)</p>	<p>none</p> <p>likely to step of line a wee bit out of line whe e final pass being just a wee bit awry in weighti k Shiny is starting a wee bit uh at the b cos this is going to be a wee bit hard to find wi It's a wee bit low (5 tokens)</p>
<p>AmE gonna be a</p> <p>only a few</p>	<p>a, 296.730 297.830 .. gonna be a dining room. 6 1055.81 I'm2] gonna be a manager of a M I think] this thing is gonna be a cold little -- 1376.77 .. it's gonna be a o=r. 1376.77 million dollars is just gonna be a drop in the bu 01.360 (H) but it's not gonna be a major change, because it's g- really gonna be a life-changing 01.78 LYNNE: And she's gonna be a ferrier. 801. 8.453 ... uh and that's gonna be a really excitin 137.834 139.413 That's gonna be a tremendous ti= or .. is there gonna be a separate, 144 20 573.210 ... there're gonna be a lot of communi 0 ... So I know there's gonna be a lot of uh - (13 tokens)</p> <p>n front of it. There are only a few seconds to dec on day 3 and J12 160 only a few on day 5 despi see. A42 95 These are only a few of the things in the Custom House for only a few morning A44 hanges J80 77 course! Only a few had the knack (14 tokens)</p>	<p>oor performance which is gonna be a very difficult This is gonna be a very interesti ricans withdraw there is gonna be a a a rather sig S1A-015(A):89 Yes it's gonna be a very interesti s it's been said there's gonna be a number of enqu (37 tokens)</p> <p>w2a-031(noone):47 Only a few species have b the product is actually only a few cards and some ough guinea worm affects only a few people in the ts dryandras, to mention only a few, are found in is posses sion tally was only a few short of his d grapple with the problem only a few years ago. (109 tokens)</p>
<p>SgE going to be a</p>	<p>has gone up and this is going to be a very excite this issue I mean is is going to be a real heavy as Prime Minister is not going to be a difficult his whole tour there was going to be a reception (34 tokens)</p>	<p>Tomorrow's going to be a hard day. and and it's not going to be <.>a ears time there's going to be a great short SEBI is going to be a statutory b (115 tokens)</p>

<p>you need a</p>	<p>You need a cold beer, fast As a Cardmember, you need a card you can rely on you know because maybe you need a medical check- length will be different you need a lot of capital get the the the software You need a software All y (25 tokens)</p>	<p>4.04 1024.89 JAMIE: Do you need a partner? 1024 so hang on when do you need a property inter But if you need a for the passpo continuous improvement, you need a long-term F logy and stuff you know you need a lot of experie (58 tokens)</p>
<p>IndE a member of</p> <p>is in a</p>	<p>his is her first term as a member of parliament the ASRC Director to be a member of the Selection country that can become a member of the stock exc n one The author is a member of the Institut h K K Gopalakrishnan was a member of the Kera (34 tokens)</p> <p>r let him say whether he is in a position to make Indeed, it is in a pickwickian sense She is in a sorrow position Yeah what is in a marriage ? (23 tokens)</p>	<p>A member of the vendors' aneda dela Rosa is also a member of our egroups a ts' needs B23 38 I am a member of the Dade Coun le Business Council and a member of the New Zeala Chinese to be elected as a member of parliament GRC one of whom must be a member of a minority gr s and I would love to be a member of the greatest , how long have you been a member of the police fo you know a roy er being a member of the the roy r r by a letter written by a member of the public t He is a member of the board alr Cora Z altho she 's not a member of our egroups. n for the high office of a Member of Parliament, s Griffiths should remain a member of my team." So and Haywood himself was a member of the escort wh ecome entitled or you're a member of the federal u (208 tokens)</p> <p>The country is in a state of euphoria he Philippine government is in a state of panic an s process and whether it is in a "material form". rotective Clothing which is in a serviceable condi cept of gracious society is in a way within the PA (68 tokens)</p>
<p>PhilE to become a</p> <p>are a lot</p>	<p>irst young man was about to become a problem, som To become a radical count She wanted to become a doctor, and s gnity it is not the way to become a people of ind ll-time faculty in order to become a president if (28 tokens)</p> <p>e thing about okay there are a lot of things I lik ut I think I think there are a lot of other eligib u mean to say that there are a lot of cut trees al</p>	<p>working class was able to become a significant p demonstrated capability to become a highly produc hat he abandon his dream to become a teacher But what must one do to become a perfect man? India has the potential to become a world leader (83 tokens)</p> <p>t and liver from animals are a lot cheaper than, p Thus, employees are a lot more involved i Yeah well there are a lot of kinds of peo</p>

	<p>neducated like uhm there are a lot of programs in (25 tokens)</p>	<p>and unfortunately there are a lot of surgeons aro (57 tokens)</p>
<p>KenE a point of</p>	<p>on. Mr. P.N. Ndwiga: On a point of order, Mr. Dep esident (Mr. Kalweo): On a point of order, Mr. Spe RITATIVE Mr. Ndicho: On a point of order, Mr. Spe Dr: Otieno-Kopiyo: On a point of order, Mr. Spe ers, I promise I'll make a point of seeing you. W (112 tokens)</p>	<p>sure it's annoying from a point of view of trying e of the peace who made a point of attending ever e Member for McKellar on a point of order S1B-0 by people who've reached a point of despair when n (56 tokens)</p>
<p>a lingua franca</p>	<p>how it began maybe uh as a lingua franca between t Thus pidgins can also be a lingua franca so long a For a language to become a lingua Franca there mus ost common name given to a lingua franca spoken th ings where Kiswahili was a lingua franca by virtue (34 tokens)</p>	<p>ch English has become a lingua franca we start the children in a lingua franca (2 tokens)</p>

Contrary to what we have seen for the definite article, clusters of *a/an* in individual varieties do not overwhelmingly contain high socio-cultural content. This is true in both tables 6.8 and 6.9. It will be shown that clusters that are found to be substantially used in individual varieties do not always demonstrate clear patterns in their concordances that can be used to explain their high frequencies. In the concluding section of this chapter, however, we will attempt to give a broad description of the possible factors affecting the patterns of these clusters as shown in the concordances. Having said this, there are several patterns or tendencies that can be identified in many of these varieties.

Two clusters in BrE, *a single currency* and *going to have a*, were analysed. The concordance shows that *a single currency* is a frequent cluster in this variety because of the socio-political developments that were taking place in Britain and its wider European context during the time of data production. Only one token was found among the other varieties. *Going to have a*, on the other hand, does not indicate any special motivation for its frequent use. It appears that the other varieties, collectively, show more varied patterns of this phrase by using more types of personal pronouns. However, BrE is still shown to use this phrase more often than any of them does individually.

In AusE, *they've got a* and *she had a* were analysed. Similar to the *going to have a* case in BrE, the concordance does not clearly reveal the motivation underlying the frequent use of *they've got a* in the Australian variety. Nevertheless, it can be observed that AusE speakers use it mostly in clause-medial position, in contrast to speakers of other

varieties who also use it in a clause-initial position. In the case of *she had a*, AusE is again shown to have a higher proportion compared to each of the other varieties. Other than this, it is difficult to ascertain the factors that drove to the high frequency of the phrase in AusE.

The first cluster from NZE, *a New Zealand*, demonstrates the kind of socio-cultural content that has so distinctly characterised many of the clusters of the definite article in individual varieties. This is the only indefinite cluster of this kind that is frequently used by a variety. With the highly specific socio-cultural content that it carries, it is not difficult to understand why the exact clause is not found in other varieties. The other cluster from NZE, *a wee bit*, is a clear example of a collective preference by speakers of a variety. The concordance shows that this phrase is used in the spoken registers as an adverbial. However in NZE it is additionally used as a non-numerical quantifier.

The two most frequent clusters from AmE, *gonna be a* and *only a few*, are not infrequent in other varieties. The clusters in AmE were thus ranked as a result of limited data being available for AmE.

SgE features *going to be a* and *you need a*. The first one offers an opportunity to compare it with the more informal *gonna be a* discussed for AmE above. It seems that the concordance does not reveal the factors underlying the high frequency of this phrase in SgE but it can be observed that the NPs used with *going to be a* in all varieties (SgE included and AmE excluded) are similar to those used with the more informal *gonna be*

a in AmE. In AmE we find *this thing, million dollars, she*, etc. as NP collocates as the phrase while for *going to be a* we find *the prime minister, the whole tour and tomorrow*. This means that the use of *gonna be a* in AmE is more varied than what we see in other varieties, i.e. *this/which/it/there's gonna be*. SgE's use of *you need a* is found to be imperative. In other varieties, the phrase is also used in questions. In spite of the fact that it is not used as a question, SgE speakers are shown to make use of this phrase more frequently than others.

In IndE, its two chosen clusters, *a member of* and *is in a*, also frequently occur in other varieties. The concordances, too, do not show any patterns that can substantively characterise their uniqueness in this variety.

Two clusters, *to become a* and *are a lot*, were analysed for PhilE. *To become a* is usually used to express a transformation into someone or something positive. In PhilE, however, it was found that this phrase is also used to express a negative transformation, i.e. *problem* and *radical counter attack*. In this variety, *are a lot* is used as part of a quantifier. In other varieties, it is also used as an adverbial. In spite of the inflexibility, PhilE is shown to use this phrase more frequently than each of the other varieties.

Finally, KenE features two clusters, *a point of* and *a lingua franca*. Regarding *a point of*, the above table shows that in KenE this phrase overwhelmingly collocates with *order*, indicating the legislative nature of its use. This finding is not anomalous if one considers that ICE-KEN contains a large amount of spoken texts sampled from courts

and the parliament. As for *a lingua franca* the majority of the tokens come from student essays. Obviously a certain topic of the composition has led the writers to write using this phrase substantially. These two cases effectively illustrate the close interrelationship between collocation and sampling.

6.4 Conclusion

The study in this chapter has demonstrated that words – grammatical words especially – co-occur with other words in systematic ways. Our aim was to exploit the collocational property of the articles to uncover their variation. First, a two-word level analysis was conducted on the definite article. Through it we discovered that some NPs such as *society*, *spring* and *church* tend to be marked for definiteness in nativised varieties. Other than *the society*, other such NPs are infrequent and not widespread. In the course of conducting this analysis, we found the claim that variation of the definite article is solely located in the nouns that collocate with it untenable. Instead we argued that it is the meaning of the article itself that leads to the variation. This occurs when it interacts with the meaning of the referent of the NP.

The analysis of R-1 collocates of *the* uncovers many types of general nouns in the varieties. We hypothesised that the use of these nouns by speakers of Outer Circle varieties would resemble that of general words by learners. We found limited evidence to support this hypothesis. A slight tendency for it to happen could be observed for IndE, SgE and KenE where spoken language was concerned. For written language, speakers of both variety groups tend to use the same number of general nouns except

for those of PhilE. Therefore when the results for spoken and written languages were combined, the Outer Circle varieties yielded only slightly more general words than the Inner Circle varieties. As for specific nouns, there were clearer differences between the two variety groups. The Outer Circle group was found to contain nouns that are used in local formal situations. Whereas these nouns are firmly positioned in Ooi's Group B, in NZE (the one Inner Circle variety where such nouns could be found), stronger efforts have been made to elevate the status of such nouns into Group A.

The R-1 analysis also revealed that each variety tends to contain a number of adjectives with high social-cultural content, although this use has not really influenced the patterns of the definite article in any substantive ways.

At the multiple-word level, the clusters analysed did not exhibit any clear and distinctive patterns of variation that can be identified along the nativeness spectrum. Instead we found numerous, though at times subtle, differences in terms of complementation or the words used to precede or succeed the clusters that can be grouped according to the following categories: Inner Circle only, antipodean only, individual varieties only. The most striking thing about the resulting groups is that none of the clusters was found to be shared only by the Outer Circle varieties or any of their sub-groups. Although these variation patterns were not found to lead to widely divergent grammatical or vocabulary patterns, it can still be argued that each variety's preferences, as has been proven for other grammatical items, has some influence on the use or non-use of the definite article. We saw this influence, for example, in the case of

the beginning of in which speakers of Inner Circle varieties use the cluster with a larger variety of prepositions as its head. The wider array of prepositions that are at these speakers' disposal allows them to choose from a greater number of noun types to construe meanings about the commencement of something. Together, the chosen preposition and noun lead to more frequent use of the cluster, and hence, the definite article.

Both analyses were repeated for the indefinite article. At the R-1 position the overlap between the varieties is small, with partitive nouns being the largest type. The amounts of general nouns in the varieties are not very different overall. AmE and KenE each has a large amount of these nouns. If the finding for AmE is discounted due to its limited and dissimilar data, the Outer Circle varieties do have a stronger tendency to use more general nouns than the Inner Circle varieties. However, on closer inspection, the use of these nouns in KenE did not really indicate the kind of generality that is associated with learner language. There was little variation observed for the L-1 collocates.

In the cluster analysis, we found that the use of this article is more homogeneous among the varieties. This was evidenced by the higher number of shared clusters between all varieties. There were some variation patterns but as in the case of the definite article, they could not be neatly divided along the Inner/Outer Circle line.

In this collocational analysis, the clearest case of nativisation has been the use of the definite article with nouns that belong in Ooi's Group B. However, these nouns

themselves do not influence the article's patterns of use in any substantive way. What is notable is that the 'noun' quality of the lexical items involved is preserved that they can successfully take the definite article and jointly form the desired NPs. In more basic sociolinguistic terms, this phenomenon would amount to 'code-mixing'.

In the collocational patterns, there is slight evidence of some form of nativisation taking place. However, because the variation patterns themselves are at times too subtle, it is hard to identify the kind of linguistic changes that have taken place. Nevertheless, one example can be offered by the *we've/he's got a* cluster. We saw above that speakers of Inner Circle varieties modify the NP complement of this cluster either with an adjectival phrase or some form of postmodification. Speakers of Outer Circle varieties, on the other hand, tend to modify it, when they do, with an adjectival phrase only. The lack/decline of postmodification of the NP complement of this cluster could be said to be evidence of 'restructuring'. However, we do need more data before the claim can be stated with more certainty.

In conclusion, the existence of these patterns does prove that there are differences in the ways different speakers 'tie' words together. Most importantly, however, if we are to uncover patterns of variation that are more robust between these two groups of varieties, we must find a different methodology than just reliance on visible collocates. We will take this up in the next chapter.

Chapter 7 The Usage Types of Articles

Collocates that are visible on the textual surface offer a lot of information about the behaviour of articles (see previous chapter). However total dependence on such collocates may mislead us into thinking that a certain usage is untypical while there is in fact a valid semantic/pragmatic reason for it (see, e.g., *the parking* in 6.1). It was pointed out that the use to which an article is put would be a better predictor of its variability. This is what this chapter aims to investigate. Section 7.1 revisits the usage types of the definite article by reviewing two seminal studies, Hawkins (1978) and Prince (1981, as well as 1992). 7.2 describe the development of annotation schemes based on the usage types of both the definite and indefinite articles. 7.3 presents the results for the definite article while 7.4 presents the results for the indefinite articles. 7.5 discusses the findings of Ø. The chapter concludes in 7.6.

7.1 Revisiting the usage types of the definite article

In chapter 6 we saw that the definite article is used with a range of NPs, which demonstrates the wide variety of usage types associated with the definite article. Many of these were discussed in section 2.1.1 (see table 2.1). While the discussion in that section was based on Quirk et al. (1985), Huddleston and Pullum (2002) and Biber et al. (1999), the present discussion explores Hawkins' usage types of the definite article (1978), on which much of the discussion in the above grammars was based, and Prince's taxonomy of givenness (1981; 1992) which enhances the insights presented in

Hawkins. These categories will serve as the basis of an annotation scheme for the definite article.

7.1.1 Hawkins' usage types of the definite article

Hawkins (1978) developed the following categories based on what was first proposed in Christophersen's (1939) familiarity theory of definiteness. In his 'location theory', Hawkins identifies the following types of definite article use: anaphoric, associative anaphoric, immediate situation uses, larger situation uses, unfamiliar uses with explanatory modifiers and uses with unexplanatory modifiers. Below, Hawkins' usage types are presented with all his examples (1978: 86-149):

Type 1: anaphoric use

This use is served by definite NPs that function as anaphors to a textual antecedent formed by another NP with the same head or one similar in meaning.

149a. Fred was discussing an interesting book in his class. I went to discuss the book with him afterwards.

149b. Bill was working at a lathe the other day. All of a sudden the machine stopped turning.

Type 2: associative anaphoric use

A slightly different kind of anaphora is formed when definite NPs are used based on the speaker's and hearer's exploitation of the knowledge they have regarding objects

evoked in the discourse and their related features or attributes. This is the associative anaphoric use. The object is called a ‘trigger’ and its feature/attribute an ‘associate’.

150. Bill drove past our house in a car. The exhaust fumes were terrible.

Type 3: immediate situation use

The immediate situation of utterance can also give rise to definite article use. Hawkins identifies two subcategories:

i. visible situation use: the referent is visible to the speaker and hearer

151. Pass me the bucket, please.

ii. immediate situation use: the referent is not visible to the speaker and hearer but can usually be inferred of its presence in the context of the immediate situation.

152. Beware of the dog.

Type 4: larger situation use

Like type 3, larger situation use is non-anaphoric but the definite NPs in this category are formed based on the hearer’s “knowledge of entities ... [in] a larger situation of utterance” (p. 115), in contrast to type 3. The size of the situation can vary but all will have the context of utterance “as their focal, defining point” (ibid.). Members of the same community typically make use of this knowledge. It is further sub-categorised into two types:

i. specific knowledge in the larger situation: an example is when the speaker and hearer are both inhabitants of a town which has a gibbet at the top of a street they both know.

Based on such knowledge, the speaker may utter (p.153).

153. *The Gibbet* [is no longer standing].

ii. general knowledge in the larger situation: this refers to “a general knowledge of the existence of certain types of objects in certain types of situation” (p. 119).

154. Have you seen *the bridesmaids*? (said at a wedding)

However, note that hearers may not depend on the same kind of knowledge to correctly locate the referent of the definite NP. *The town clerk*, for example, can be located correctly by a hearer either by virtue of specific knowledge (e.g. acquaintance with the individual) or general knowledge, i.e. a town typically has a town clerk.

Type 5: ‘unfamiliar’ uses in NPs with explanatory modifiers

Definite NPs that do not belong to the above categories are those that Hawkins classifies as ‘unfamiliar’ due to their obvious oppositeness to the familiarity theory.

There are four types of explanatory modifiers that form this category, listed below. To account for the felicitousness of *the*, Hawkins applies transformational rules to the relevant clauses.

i. establishing relative clauses

Locatability of a referent is based on the content of the relative clause. Without previous mention (thus defying familiarity), these clauses “relate the new, unknown object either to other objects in the previous discourse set, or to participants in the speech act, or else they identify entities in the immediate situation of utterance” (p. 137).

155. What's wrong with Bill? Oh, the woman he went out with last night was nasty to him.

ii. associative clauses

These clauses incorporate both the trigger and the associate found in a sequence indicating an associative relationship between two objects.

156. I remember the beginning of the war very well.

iii. NP complements

These are first mention uses of definite NPs whose heads are modified by a complement.

157. Bill is amazed by the fact that there is so much life on Earth.

iv. nominal modifiers

These are first mention NPs made definite by post nominal modifiers.

158. I don't like the colour red.

Type 6: unfamiliar uses with 'unexplanatory' modifiers

These are definite NPs formed by a small set of words. According to Hawkins, they are unexplanatory based on the fact that they do not function to introduce the unknown to the hearer.

159. My wife and I share the same secret.

A different, but related, way of conceptualising the notion of identifiability is to make explicit the role of the hearer and the speaker in discourse. Information that is relayed from the speaker to the hearer can be classified as ‘old/given’ or ‘new’ based on this configuration. The following section describes Prince’s attempt at elaborating these relationships in detail.

7.1.2 Prince’s taxonomy of givenness

NPs are old/given if they have been introduced to the hearer and speakers would mark these NPs for definiteness accordingly. This is in fact the basic idea that we have seen in Hawkins’ categories. However, for Prince, newness should not only be defined in terms of the hearer (1981; 1992). If an entity is introduced into a discourse for the first time, the NP is then said to be discourse-new. And this is regardless whether it is hearer-new or not. Nevertheless, when an NP is hearer-new, by definition, it is discourse-new too. Similarly, a discourse-old NP entails a hearer-old one. Further relationships between old/given and new entities in terms of hearer and discourse are described in the following taxonomy. All examples are from Prince.

a. Brand-new

Hearer-new and discourse-new entities are usually introduced by indefinites (though, of course, not necessarily articles). These are called ‘brand-new’ by Prince.

160. I’m waiting for it to be noon so that I can call *someone in California*.

b. Brand new anchored

Prince claims that a brand-new entity is anchored if there is some kind of link that exists between itself and another entity in the discourse. This link provides information that is not new to the hearer.

161. A guy I work with says he knows your sister.

c. Evoked

Evoked NPs are of two types: ‘textually evoked’ and ‘situationally evoked’. Entities can be found to be already established in the discourse-model, i.e. evoked, by virtue of the discourse itself or the situation of utterance. (162) may be both hearer-old and discourse-old. (163), on the other hand, may be hearer-new and discourse-new (cf. Hawkins’ *pass me the bucket* above).

162. A guy I work with says he knows your sister (textually evoked).

163. Pardon, would you have change of a quarter? (situationally evoked)

d. Unused

When hearer-old but discourse-new entities are evoked, they are called ‘unused’. The entities have been introduced to the speaker/hearer but up to the point of utterance in the discourse, they have not been mentioned, i.e. used.

164. I’m waiting for it to be noon so that I can call Sandy Thompson.

e. Inferrables

Discourse entities which can be inferred through their association with other evoked

entities are inferrables (cf. Hawkins' *wedding – the bridesmaids* above). Prince claims that inferrables can be inferred via their association with other inferrables too. In Prince's terms, however, inferrables are neither hearer-old nor discourse-old. Despite this, they are not hearer-new or discourse-new either.

165. I got on a bus today and *the driver* was drunk.

f. Containing inferrables

An inferrable may actually be present as part of the NP itself, i.e. 'contained' in it.

166. *The door of the Bastille* was painted purple.

Used in conjunction with familiarity (in the sense explicated by Hawkins' location theory), discourse-newness is a useful concept when discussing certain definite article usage patterns found in the Outer Circle. Presently we will try to exploit Hawkins' and Prince' insights and establish categories for the purposes of our study.

7.2 Annotation schemes

This section will describe the development of two annotation schemes, one for the definite article and one for the indefinite article.

7.2.1 Developing an annotation scheme for *the*

Poesio and Vieira's (1998) annotation scheme is also based on Hawkins and Prince. However, the findings of their experiments with corpus annotation of definite NPs caution us to minimise the number of coding categories to be used. In their study it was

found that categories that follow too closely those of Hawkins and Prince are not thoroughly reliable since overlap in the categories and simply having too many categories can lead to confusion for the annotator. In addition, Hawkins' categories have been criticised for being prone to ambiguity (e.g. Fraurud, 1990).

Based on Hawkins' categories, a reduced set is proposed by Liu and Gleason (2002). Their study is intended to investigate ESL students' acquisition of the definite article and to ascertain if its different non-generic uses "are acquired at the same time" (p. 1). The creation of Liu and Gleason's four categories from Hawkins' is explained in figure 7.1:

Figure 7.1: Correspondence between Liu and Gleason's and Hawkins' categories

Liu and Gleason	Hawkins
Textual	anaphoric (type 1), associative anaphoric (type 2)
Situation	immediate situation use (types 3i and ii), larger situation use (type 4i)
Cultural	larger situation use (type 4ii)
Structural	unfamiliar uses with explanatory modifiers (type 5), unfamiliar uses with unexplanatory modifiers (type 6)

Although Liu and Gleason claim that the above categories helped to explain the phenomenon they were investigating (2002: 8), there are a few arguments that can be

made against their adoption in this study. First, the reliance on the term ‘culture’ in the cultural category raises more questions than those that it tries to answer. According to Liu and Gleason:

Although [Hawkins] included what we call cultural use in the situation category, we believe it should be a separate type because, although such use shares some characteristics with situation use, it differs from the latter in two important ways. First, whereas situation use [types 3i and 3ii] and larger situation use relying on specific knowledge [type 4i] refers to instances where the referent is within the view of the interlocutors or is known to everyone in a community, cultural use [type 4ii] does not have that luxury. Instead the interlocutors have to assume and resort to a shared knowledge in the entire language community to make such use functional. Second and more importantly, such use of *the* is often not framed by situation but is determined, to a large extent, by conventional practice.

(2002: 7)

In the above description, ‘culture’ seems to refer to:

- (1) a shared knowledge in an entire language community;
- (2) conventional practice.

The problem with (1) is that there is no sufficient definition given for *shared knowledge in a language community* so as to differentiate it from Hawkins’ type 4ii. The cultural use, they argue, can be exemplified in the use of *the* in some river names but not in those of most lakes. What confounds this argument is their simultaneous claim that “the use of *the* in [such] circumstances is not completely arbitrary but also governed by rules” (p. 7). If indeed such use is rule-governed, then there should not arise any need to invoke culture as its basis. In fact when we probe Liu and Gleason’s argument further, it appears that the term ‘culture’ is actually used to describe knowledge that is more widely shared by members of a community. This is despite their contention that “interlocutors have to assume and resort to” a different kind of knowledge in a

language community (p. 7). But it is not clear what this knowledge entails and the difference between a *community* and a *language community* is just as vague. Thus, Hawkins' type 4ii still captures best the knowledge needed here. The term 'culture' is clearly too nebulous to offer anything concrete about knowledge that is shared by members of a language community. This brings us to the problem with (2) or conventional practice. According to the authors, cultural use of *the* depends on conventional practice. However, based on the investigation of the definite article that has been carried out thus far, culture does not always guarantee conventional practice. Consider the use of both 'Ohio University' and 'the Ohio University' by speakers of AmE (see 94 and 95 in 6.1.1). Because this variability is found among speakers of the same culture, the choice between the definite article and the null one cannot be solely attributed to this broad culturally based conception of shared knowledge. While culture is no doubt present in the situation in which the choice is made, the decision seems to be more plausibly related to the interaction between circumstantial factors such as the speaker's familiarity with the referent and the linguistic representation needed to encode it. This formulation leaves culture out but retains Hawkins' 'larger situation' in the definition. Furthermore, Liu and Gleason's other examples of this category raise serious doubts about the tenability of the 'culture' construct that they rely on. As evidence of conventional practice, they list musical instruments such as *the piano* in 'play the piano' (p. 7) and common illnesses such as *the flu* in 'I've got the flu' (p. 20, fn. 7). Although they rightly claim that speakers of English typically say 'the piano', 'the guitar', 'the flu', 'the plague' etc. but not 'the basketball', 'the cricket', 'the polio', 'the cancer' etc., these speakers' knowledge of how to select *the* appropriately in such

cases does not primarily stem from convention but, as is clear from its meaning, from the desire to express these NPs as generic. Thus they are irrelevant to the category being discussed. It can then be concluded that Hawkins' use of the term 'situational' is more appropriate.

There is also a problem with the textual category as proposed by Liu and Gleason. It seems rather simplistic to treat anaphoric and associative anaphoric equally by categorising both of them as textual. While both categories have a textual trigger, the association between the trigger and the referent in associative anaphoric depends on knowledge that is external to the text. It is this knowledge that determines the success of the determination (Prince, 1981; 1992, Clark 1977, and Sidner 1978). It would be incorrect to suggest, as Liu and Gleason do, that the use of *the* of this kind is entirely textual in nature. To illustrate the centrality of this non-textual knowledge in ensuring the felicity of an associative anaphoric definite, consider the following invented discourse regarding the Singaporean traditional practice of giving 'ang pows':

167. Leonard realised that he had accidentally taken his impertinent brother's ang pow. *The packet* had been opened. His heart sank when he saw how much he received from Auntie this year.

The listener/reader of these sentences is required to know that 'ang pow' is money given in an envelope (or packet), especially during the celebration of Chinese New Year.⁴⁸ This knowledge must be available to him/her so that the trigger *ang pow* in (167) can lead to its association with *the packet*. For someone unfamiliar with the

⁴⁸ Children are the ones who most commonly receive 'ang pows' and it is generally considered rude to open 'ang pows' in the presence of those who give them.

traditions of Singaporean or other similar Asian societies, it is not hard to imagine the difficulty s/he is faced with in interpreting this NP.⁴⁹ There are of course countless other examples that can illustrate the insufficiency of the text alone in the interpretation of associative anaphoric definites. The important point is that the use of *the* in the associative anaphoric category cannot be said to be textually motivated only. Taking into consideration the role that non-textual knowledge plays in the interpretation of this kind of NPs, Hawkins' associative anaphoric category (or Prince's inferrable) must be preserved.

It remains necessary for our proposed categories to be fewer than Hawkins' and Prince's. In reducing these categories, some of them will be collapsed based on their textual and semantic properties, while others will be merged through their syntactic and/or lexical properties (see table 7.1). Prince's categories, incorporating discourse new/old notions, complement Hawkins' and facilitate the reduction task. Poesio and Vieira (1998) claim that not all of Prince's classifications may correspond with Hawkins'. These incompatibilities are sorted out below:

a. Hawkins' *unfamiliar uses* vs. Prince's *containing inferrables*: Hawkins' unfamiliar uses are grouped together based explicitly on their syntactic structures. These include NP-complements, nominal modifiers, establishing relative clauses and associative clauses. One type, however, is placed in this group based mainly on its lexical property

⁴⁹ It is interesting that Liu and Gleason do not consider associative anaphoric definites under their cultural category, which at some point would be more relevant to their 'culture' construct.

(e.g. *the same secret*). If we rely on Prince's intuition that these types are also inferrables "where what is inferred off of is properly contained within the Inferrable NP itself" (1981: 236), then it follows that this 'inference within inference' is necessarily syntactically marked, and is manifest in precisely the syntactic forms identified by Hawkins. The use of *the* in this category is clearly structurally based. As for *unexplanatory modifier use*, for convenience, we will regard this small set of modifiers as consistently requiring a particular syntactic structure to form an NP, i.e. 'the + modifier + noun [+ mod phrase]', due to their unique lexical properties. This class of definites is called 'structural'.

b. *associative anaphoric uses vs. inferrables*: this class of definites is known to be complex (Prince, 1981: 236) and there have been many attempts to account for its working mechanism (see Clark, 1977; Sidner 1978). It does not fit easily in either the textual or situational category. While the associative anaphoric antecedent is textually established, the 'coreferent' (Sidner, 1978) is arrived at through a variety of non-textual relationships (see 167 above). These relationships need not concern us here (see Clark, 1977 and Sidner, 1978), but the intricacy between them must be recognised and a separate category for them has to be created (or preserved). This class will be known by Hawkins' term for it i.e. associative anaphoric.

c. Following Chesterman (1991: 53), two more uses are added: 'non-referential' and 'generic'. The first use is exemplified by 'Chang is *the leader*', and the second use by '*The computer* is a wonderful thing'. A third added category is 'idiomatic', created to

capture 'frozen' expressions such as 'shoot *the breeze*' (cf. Poesio and Vieira 1998). Thus, the categories to be used in this study consist of seven types i.e. textual, structural, situational, associative anaphoric, non-referential, generic and idiomatic, although in the actual coding there will also be instances of 'unknown', 'repeat' and 'repair'. The formation of these categories from Hawkins (1978) and Prince (1981) is summarised in table 7.1.

Table 7.1: Correspondence between Hawkins', Prince's and the present study's categories

Hawkins	Prince	Categories in the study
<p>Anaphoric use</p> <p>Immediate situation use <i>i. visible situation use:</i> <i>ii. immediate situation use:</i></p> <p>Larger situation use <i>i. specific knowledge in the larger situation:</i> <i>ii. general knowledge in the larger situation:</i></p> <p>Associative anaphoric use</p> <p>'Unfamiliar' uses in NPs with explanatory modifiers <i>i. establishing relative clauses:</i> <i>ii. associative clauses:</i> <i>iii. NP complements:</i> <i>iv. nominal modifiers:</i> <i>v. 'unexplanatory' modifiers</i></p>	<p>Textually evoked</p> <p>Situationally evoked Unused Inferrable</p> <p>Inferrable</p> <p>Containing inferrable</p>	<p>Textual</p> <p>Situational</p> <p>Associative anaphoric</p> <p>Structural</p> <p>Non referential <i>e.g. He is the leader.</i></p> <p>Generic <i>e.g. The computer is a wonderful thing.</i></p> <p>Idiomatic <i>e.g. Grab the bull by its horns.</i></p> <p>Unknown <i>e.g. the guy (with insufficient context for identification)</i></p> <p>Repair <i>e.g. The, er, the project is important for several reasons.</i></p> <p>Repeat <i>e.g. The, the man came early today.</i></p>

7.2.2 Developing an annotation scheme for *a/an*

For the indefinite article *a/an*, the annotation scheme was mainly based on insights from three major grammars, Huddleston and Pullum (2002), Biber et al. (1999) and Quirk et al. (1985) (see table 2.1). A few modifications that had to be made are described below.

The difficulty to distinguish between first-mention indefinites and quantitative indefinites, which occurs quite frequently, is illustrated in (168):

168. Q: >um<> sara's got a really nice sleeveless green you know
coat jacket

B: >yeah<><mm it's really great

<ICE-NZ:S1A-002#221:1-224>

As a result these categories are merged bearing in mind that in some cases one meaning is more prominent than the other.

The non-referential use of the indefinite article following a copulative verb (e.g. 'He is *a liar*') has to be differentiated from the non-referential use of a generic indefinite as in:

169. Men may initially have to face the fear that they may be found to be infertile - -
which many think makes them less of a man <ICE-SIN:W2C-002#79>.

The solution is to establish a category called 'factive', a term originally used by Sankoff and Mazzie (1991) to classify what they claim as non-specific nouns. However Sharma observes that these NPs are predicative and "perform neither a strictly specific

function nor an entirely non-specific function, but rather mark category-membership” (2005: 547). The category is thus used in lieu of the broader ‘non-referential’. Figure 7.2 lists the categories the resulting annotation scheme.

Figure 7.2: Categories for *a/an*

first-mention/quantitative
factive
generic
idiomatic
unknown
repeat

7.2.3. Analysing Ø

The zero article, as well as its null counterpart, presents the greatest challenge to the corpus-based methodology adopted in this study. To analyse the patterns of Ø, a case study of the noun ‘house’ was conducted. Tokens that function as verbs or adjectives or form part of titles of books, films, etc. (e.g. ‘The House of the Rising Sun’ or ‘Little House on the Prairie’) were excluded from the analysis. Due to problems in providing equivalent data for AmE (as described methodologically in chapter four and in the results reported in chapters five and six), analysis of this variety was not made.

The annotation scheme consists of two usage types, i.e. determined and undetermined. Undetermined tokens of ‘house’ was subjected to further qualitative analysis.

7.2.4 Data and procedures

Due to practicality considerations, not all text types or “genres” (Baker et al., 2006: 77) were analysed. The analysis presented below is based on three text types from each corpus, i.e. private dialogue, academic writing and reportage. Their selection was based on the need for variation; these are known to be different genres (e.g. Biber et al., 1999). The number of NPs analysed varies according the article being investigated. Between the two types of articles, more NPs were analysed for the definite article due to the simple fact that this article is more frequently used in all categories.

For the definite article, approximately 500 NPs were targeted for each category. In private dialogue, to ensure that approximately 500 NPs were tagged, seven files in this category from each corpus were analysed. These were the first seven files, sorted according to file number. As for academic writing, one file, i.e. the first one, was selected from each of the subcategories of ‘humanities’, ‘social sciences’, ‘natural sciences’ and ‘technology’. The four files only yielded 400 – 700 tokens from each of the corpora. As regards reportage, the analysis was performed on the first four files, sorted according to file number, which also gave 400 – 700 tokens from each corpus. As far as the definite article is concerned, approximately 1500 NPs from each corpus or variety were tagged, although tokens of unknowns, repeats and repairs were subsequently discarded.

As regards *a/an*, there were much fewer tokens of this article in each of the categories. Using the selection criteria for the definite article, 150 – 250 tokens were analysed for

private dialogue, 150 – 350 tokens for academic writing and 140 – 250 tokens for reportage.

Private dialogue in AmE was represented by four files from SBC. These files yielded 586 tokens of *the* and 548 tokens of *a/an*. Academic writing was represented by four files from the learned and scientific category in FROWN. The files were selected from different academic fields to parallel the selection of the four files in ICE's academic writing category. 190 tokens of *a/an* were analysed from these files. Finally, for AmE, reportage consisted of four files from FROWN's corresponding category, also called reportage. A total of 237 tokens were analysed based on these files.

Each token of *the* or *a/an* was coded according to the categories described in table 7.1 and figure 7.2 respectively. The coded tokens from each variety or category are described in terms of percentage. The chi-squared test was used to assess the significance of any relationships observed. Tokens from Outer Circle data that did not display the typicality that is otherwise demonstrated by the majority of their Inner Circle counterparts for a particular category were analysed further. It must be added that some Inner Circle tokens are also less typical than others; these too were subjected to further analysis.

7.2.5 Sample cases and problematic cases for the definite article

In this section we describe how each category in the definite article annotation scheme is delimited and some inevitable problems that can be found in some of them. It must be pointed out that tokens that straddle two or more categories are an indication that these categories are, in the first place, an attempt at simplifying what really are an abstract and fuzzy phenomenon (see Lakoff, 1982 on linguistic categorisation from a cognitive perspective and Rosch et al., 1976 on prototypicality effects). Categorisation, however, does help to make sense of linguistic structures in terms of prototypicality (Huddleston and Pullum, 2005: 22).

The essential criterion for the textual category is an antecedent. Not only does the antecedent have to be realised textually, but it also must share the same head or a semantically similar one with the definite NP. How far apart can an antecedent be from its anaphor before the co-reference is dissolved? Based on measurements of cognitive accessibility of NPs by discourse participants, Givon proposes a maximum distance of seven clauses for the antecedence to be preserved if it is established by a definite NP and a maximum of ten clauses if it is by a modified definite NP (1992:21). These are the limits that are observed in this study.

The structural category derives its name from its dependence on certain syntactic structures to modify the nominal and render it identifiable. Most prenominal modifications, such as ‘the big bus’, are not categorised this way. In the data, the most common syntactic pattern to be used is postnominal modification via relative clauses,

prepositional phrases and certain adverbs. As explained in 7.2.1, certain lexical items, mostly adjectives, have in them properties that make the nominals they modify consistently definite. The resulting syntactic pattern is highly regular and thus conceived this way: ‘the + lexical item + noun’. Hawkins’ *the same secret* is an example. Others include *first, last, following* and superlatives. *Rest* is one of few nouns that constantly appear in the company of *the*.

In contrast to the previous two categories where they seem to have an internal ‘source’ or ‘provider’ of definiteness in the textual sense, the situational category, as its name suggests, captures many definite referents which are formed based on the states of affairs in the broader non-textual context of utterance. It has a very important implication for the study of *the* in the varieties. It allows us to think this way: ‘when a definite article is used it is a marker of identifiability of some kind; therefore there is something in its situational context that can explain how this referent is identifiable’.⁵⁰

Here is an example:

170. B: Next time I’ll bring you to the restaurant at Sogo. They have *the*

Japanese pancakes

A: Is it the one just outside Sogo

B: No it’s at Paragon you know

<ICE-SIN:S1A-013#52:1 – 55>

⁵⁰ This is also the assumption for *unknown* definite NPs, but based on the pragmatic knowledge of the annotator, the referent cannot be identified.

There is no mention of *pancakes* prior to this utterance and no suggestion in the context that *pancakes* is situationally evoked. One way of accounting for the pancakes is to think of a motive that B might have when uttering it. It could be that B thinks that A knows which pancakes she wants to talk about and she wants A to locate this referent (see Hawkins, 1978). Whatever it is, we would normally expect people (native speakers, at least) to say “Next time I’ll bring you to Sogo. They have \emptyset *Japanese pancakes*”.⁵¹ This brings us to the issue of typicality. If this pattern is accepted in an L2 variety, we will find relatively frequent occurrences of such an NP to confirm it.

The associative anaphoric category captures a definite reference triggered by a textual element but established through a variety of non-textual factors. Often these involve cultural elements (but see 7.2.1). As argued above, the involvement of non-native cultures could pose a problem to the annotator if relationships that are otherwise the norm in a particular culture are unknown to him or her (see 167). Consider also this example:

171. Uhm uhm at this moment the COE is so expensive then the car the second-hand car prices will go up also <ICE-SIN:S1A-013#134:1>.

⁵¹ Most probably the speaker is mentioning *pancakes* for the first time here. Or it may also be the case that there is nothing special and thus identifiable about these pancakes other than their being called Japanese. Thus, is it possible that *the Japanese pancakes* is actually used generically here? Sidner (1978: 90) claims that “generics in the midst of a discourse can be easily disambiguated, but in an initial sentence only implicit contexts may exist in which a co-referent can be found to use in disambiguation”. It appears that Outer Circle uses of *the* suggesting genericity e.g. *the pancakes* cannot be easily submitted to this claim (i.e. it appears mid-discourse). It has been argued that *the* + plural NP structure too can give a generic interpretation (Burton-Roberts, 1976; Quirk *et al.*, 1985). Chesterman notes that in the example ‘Among the lizards, iguanas are the most popular as a local food, the lizards’ indicates “types, subspecies of a genus” (1991: 36). *The Japanese pancakes*, on the other hand, is only one type of pancakes, so this interpretation does not hold. It seems that the use of the definite article delimits the qualitative generality (extensivity), rendering *pancakes* specific and identifiable: a subset of pancakes known to the speaker and presumably to the hearer too (see 2.5.3).

The trigger of *the second-hand car prices* is the COE (certificate of entitlement) which is actually an amount of money paid to transport authorities as evidence of one's financial ability to purchase and maintain a vehicle in Singapore. (Note that *the car* before *the second-hand* is a repair.) In the event that such a culture-specific relationship cannot be confirmed, a potential associative anaphoric use is considered 'unknown'.

A token can also be unknown due to transcription. In the following, *the first batch* can either be structural or factive. This is due to the transcription of *there're* before it, which can be interpreted as 'there are' or 'they are'. It is also based on the nearly identical pronunciation of the contracted form of these phrases in SgE.

172. And then there're there're *the first batch* that reached there <ICE-SIN:S1A-001#148>.

Finally, a token can be unknown because its category is unknown to the annotator (cf. *the second-hand car prices* in 171).

The repeat and repair categories are quite straightforward and they are included due to the usually on-line nature of spoken language. However, preliminary analyses indicated that these categories are not central to the meanings associated with definite article use (Wahid, 2007 and 2008) and are thus not discussed any further in this thesis.

The generic category is made of obvious generic reference such as:

173. MARY: But I think running it out on *the open road*, will cause it ... possibly to shoot more <SBC 910.76 – 914.72>.

Nevertheless generic reference is not always this clear cut. Genericity is considered as reference to a kind or a generalised description (Carlson and Pelletier, 1995: 2), such as *the open road* in the above. At the same time, there are tokens of *the* that seem to incorporate characteristics of generic and situational categories. In the following excerpt where several siblings, some of whom are teachers, are discussing how they treat their students in class, notice how *the table* seems to, on the one hand, behave like *the store* in Birner and Ward's 'He's gone to *the store*' (1994) where *the store* can refer to any store. In this sense the definite NP is situational because, despite the ambiguity (cf. Du Bois, 1980), the hearer is willing to locate the store as one of the many that exist in the vicinity. However, in this case, *the table* also seems to represent the subset of tables that belonged to all students in the class, thus suggesting kind in a limited sense. The fact that it is used in the singular also points to the direction of this NP being generic.

174. SHARON: You're too nice, you're too nice, your kids are gonna take advantage of you. And they were. Th- th[ey k-] --

KATHY: [Yeah].

SHARON: They came to where, they'd be talking the whole time, during the as[sign]ment,

CAROLYN: Ah.

SHARON: I'd be up at the board, they'd be, I'd turn [around they'd go],

CAROLYN: ah

SHARON: hee hee hee, cause they didn't think I was serious.

SHARON: [And I] finally got to the point where, you know, every day, I'd have em put their heads on *the table* for ten minutes[.]

<SBC 672.45 – SBC 690.58>

On the use of *the* as a limited genericity marker, Chesterman observes:

the indicates 'all (members) of a locatable set'. This basic meaning remains constant, but the extension of the set may vary. At its most limited, the extension may be no more than a single individual particular – a one-member set – or a group of particulars. But in generic contexts the extension expands ... in the case of a one-member set the one member may be a subset rather than an individual[.]

(1991: 75)

Chesterman's illustrates this with *the pony* in 'The pony found in Hampshire is the New Forest pony' and contrasts it with generic cases where the extension is "maximum", e.g. 'The otter is a dying species' (ibid). It appears that *the table* above is of the former type, which gives it, and other similar NPs, a generic categorisation in this study.

The idiom category in Poesio and Vieira (1998) consists of indirect references, idiomatic expressions and metaphorical uses (p. 192). In this study, Grant and Bauer's (2004) test for core idioms is utilised to qualify NPs for this category. By their criteria only "multiword units" such as 'kick the bucket' or 'shoot the breeze' are true idioms. This is because such phrases are non-compositional and cannot be "reinterpreted pragmatically to understand the intended truth" (ibid: 51). The NP *the game* in (175) below is a figurative by the same criteria and thus considered situational in this study.

The fact that *the game* can occur in both *rules of the game* and *play the game* already indicates it is not an idiom in which the intended meaning stays constant. Instead it is a figurative, which in fact allows a degree of variability.

175. DARRYL: yeah but .. but .. but, to me the whole point is is, you have no idea, what happens before or after. You have no idea. You can read books about it, and you can .. talk about it, but the most pragmatic thing to do is, to just ... live it.

PAMELA: Hm.

DARRYL: Learn the rules of *the game*, play *the game*[.]

<SBC 619.12 – SBC 646.32>

Finally, we need to address the inherent ambiguity of Hawkins' (and by the same token, Prince's) categories that has been criticised by Fraurud in her 1990 study. This concerns the claim that definiteness as encoded by *the* can originate from different sources.

In some cases, the context of corpus data does allow the analyst to make safe guesses about what kind of definite NP is being made. In some other cases, this is less clear. For example, on finding *the Prime Minister* in a British newspaper, a reader could identify this entity based either on his knowledge of the British culture or an antecedent in the text. Fraurud observes:

[t]he identification of [triggers] could possibly be more adequately described as a simultaneous processing of different kinds of available information ... there would appear to be an interaction between lexico-encyclopaedic knowledge

associated with the head noun of the definite NP and the general knowledge associated with currently present anchors” (ibid., 426).

It would appear that the task of describing this interaction is currently impossible for the analyst given the ‘secondary’ nature of corpus data. However, this is precisely the position of a corpus analyst: s/he cannot extrapolate from the data that s/he does not have. But whatever data that are available to her/him can be exploited in any useful way. This is where categories can be of assistance. They help position the analyst’s interpretation clearly and in turn suggest the nature of “the lexico-encyclopaedic knowledge associated with the head noun of the definite NP and the general knowledge associated with currently present anchors” (ibid.). More importantly, if the categories are to classify the data at all, it must be because the categories allow the analyst to do so. In regard to the usage of *the*, the categories have to be set up in a way that there is no overlap between them (but see prototypicality effects above). In the present analysis, each definite NP that seems to qualify for several categories is analysed in the ways described below. These formulas are not based on arbitrary choices. Rather their selection is based on what appears to be their salience, e.g. structural over associative anaphoric and situational in formula (A).

A: If an NP is eligible for associative anaphoric or situational or structural category, it will be classified as structural. Below is an example.

176. A: I fly by Airways to Bangkok

C: And then

A: And then uh coach up to Chengmai

C: Oh

A: The rest would be coach ride overnight
C: And then back to Singapore
A: Fly from Bangkok
C: Orh that's not so bad. But the trip from Bangkok to Chiangmai. How long does that take

<ICE-SIN:S1A-003#111-119>

In B, *the trip* is triggered by a previous mention of *fly, Bangkok, coach up to Chengmai*, etc., making it a candidate for the associative anaphoric category. However, here it is categorised as syntactically patterned.

B: If an NP is associative anaphoric, situational or structural and repeated in whole as a discourse referent, then the second token of this NP is classified textual. Below is an example.

177. SHARON: Well they really are picking on the fact that I'm new, like, y- --
uh, It's really annoying.

CAROLYN: The fact that you're new, uh, uh,

SHARON: That I

CAROLYN: I would, I would go further than that.

<SBC 492.52 – SBC 497.83>

In B, although *the fact that you're new* is originally structural, it is repeated in whole by Carolyn as a response to Sharon. The NP in Carolyn's response is textual.

7.3 Results: *the*

A total of 14,073 tokens of *the* were analysed in the corpora. 7,123 were from the Inner Circle group and 6,950 were from the Outer Circle. Table 7.2 shows the breakdown of the analysed tokens according to varieties and categories.

According to the table, the definite article was used most frequently either in the situational or structural category. This is in accordance with Fraurud's contention that the definite article is not most prominently used for its textual function despite being commonly described as such (1990). The textual category itself is found in the third place in all the varieties. Factive is the fourth placed category for the majority of the varieties. The next two places are occupied interchangeably by the generic and associative anaphoric categories. Idiomatic is placed last in all varieties.

Table 7.2: Tokens of *the* according to varieties and categories

		Textual	Structural	Situational	Factive	Idiomatic	Generic	A. Anaphoric	Total
AusE	Priv. Convrs.	77 (17%)	93 (21%)	187 (42%)	28 (6%)	7 (2%)	19 (4%)	34 (8%)	445 (100%)
	Acad. Writg.	79 (14%)	223 (38%)	264 (45%)	6 (1%)	0 (0%)	4 (0.7%)	11 (2%)	587 (100%)
	Reportage	98 (16%)	158 (26%)	328 (54%)	11 (2%)	1 (0.2%)	2 (0.3%)	3 (0.5%)	601 (100%)
	Total	254 (16%)	474 (29%)	779 (48%)	45 (3%)	8 (0.5%)	25 (2%)	48 (3%)	1633 (100%)
BrE	Priv. Convrs.	59 (14%)	172 (40%)	147 (34%)	13 (3%)	2 (0.5%)	17 (4%)	18 (4%)	428 (100%)
	Acad. Writg.	93 (15%)	285 (46%)	206 (33%)	4 (0.6%)	7 (1%)	2 (0.3%)	20 (3%)	617 (100%)
	Reportage	76 (14%)	214 (38%)	237 (42%)	19 (3%)	1 (0.2%)	4 (0.7%)	13 (2%)	564 (100%)
	Total	228 (14%)	671 (42%)	590 (37%)	36 (2%)	10 (0.6%)	23 (1%)	51 (3%)	1609 (100%)
NZE	Priv. Convrs.	79 (19%)	118 (28%)	174 (41%)	5 (1%)	4 (0.9%)	22 (5%)	24 (6%)	426 (100%)
	Acad. Writg.	89 (9%)	520 (52%)	324 (33%)	16 (2%)	12 (1%)	17 (2%)	20 (2%)	998 (100%)
	Reportage	129 (22%)	156 (27%)	272 (47%)	5 (0.9%)	1 (0.2%)	8 (1%)	7 (1%)	578 (100%)
	Total	297 (15%)	794 (40%)	770 (39%)	26 (1%)	17 (0.9%)	47 (2%)	51 (3%)	2002 (100%)
AmE	Priv. Convrs.	204 (24%)	184 (22%)	263 (31%)	80 (10%)	0 (0%)	15 (2%)	89 (11%)	835 (100%)
	Acad. Writg.	52 (10%)	276 (53%)	178 (34%)	7 (1%)	3 (0.6%)	0 (0%)	7 (1%)	526 (100%)
	Reportage	128 (25%)	112 (22%)	251 (49%)	13 (3%)	4 (0.8%)	6 (1%)	4 (0.8%)	518 (100%)
	Total	384 (20%)	572 (30%)	692 (37%)	100 (5%)	7 (0.4%)	21 (1%)	100 (5%)	1879 (100%)

SgE	Priv. Convr.	76 (19%)	83 (21%)	210 (53%)	9 (2%)	0 (0%)	8 (2%)	14 (4%)	400 (100%)
	Acad. Writg.	230 (32%)	237 (33%)	207 (28%)	10 (1%)	9 (1%)	11 (2%)	27 (4%)	730 (100%)
	Reportage	150 (30%)	144 (29%)	140 (28%)	5 (1%)	8 (2%)	8 (2%)	45 (9%)	500 (100%)
	Total	456 (28%)	464 (29%)	557 (34%)	24 (2%)	17 (1%)	27 (2%)	36 (5%)	1630 (100%)
PhilE	Priv. Convr.	101 (20%)	133 (27%)	176 (35%)	42 (8%)	6 (1%)	11 (2%)	31 (6%)	500 (100%)
	Acad. Writg.	122 (16%)	332 (42%)	259 (33%)	11 (1%)	7 (1%)	31 (4%)	27 (3%)	789 (100%)
	Reportage	186 (22%)	211 (25%)	404 (49%)	7 (0.8%)	3 (0.4%)	1 (0.1%)	18 (2%)	831 (100%)
	Total	409 (19%)	676 (32%)	839 (40%)	60 (3%)	16 (0.8%)	43 (2%)	76 (4%)	2120 (100%)
IndE	Priv. Convr.	77 (19%)	140 (35%)	146 (37%)	15 (4%)	1 (0.3%)	9 (2%)	9 (2%)	397 (100%)
	Acad. Writg.	89 (16%)	307 (54%)	128 (23%)	2 (0.4%)	5 (0.9%)	1 (0.2%)	34 (6%)	568 (100%)
	Reportage	141 (25%)	141 (25%)	233 (41%)	5 (0.9%)	7 (1%)	4 (0.7%)	33 (6%)	564 (100%)
	Total	307 (20%)	588 (39%)	507 (33%)	22 (14%)	13 (0.9%)	14 (0.9%)	76 (5%)	1529 (100%)
KenE	Priv. Convr.	124 (28%)	104 (24%)	125 (29%)	14 (3%)	45 (10%)	12 (3%)	13 (3%)	437 (100%)
	Acad. Writg.	119 (18%)	261 (40%)	204 (31%)	14 (2%)	6 (0.9%)	34 (5%)	20 (3%)	660 (100%)
	Reportage	128 (22%)	177 (31%)	236 (41%)	15 (3%)	1 (0.1%)	5 (0.9%)	11 (2%)	574 (100%)
	Total	371 (22%)	542 (32%)	565 (34%)	43 (3%)	52 (3%)	51 (3%)	44 (3%)	1671 (100%)
Total		2706 (19%)	4781 (34%)	5299 (38%)	356 (3%)	140 (1%)	251 (2%)	532 (4%)	14073 (100%)

In order to determine if the varieties use significantly more or fewer tokens in certain categories than others, we need to find out the extent to which each category is constrained by the varieties. Table 7.3 shows the comparison in this perspective while figure 7.3 is a corresponding series of bar charts illustrating the differences.

Table 7.3: Percentages of each category of *the* according to varieties

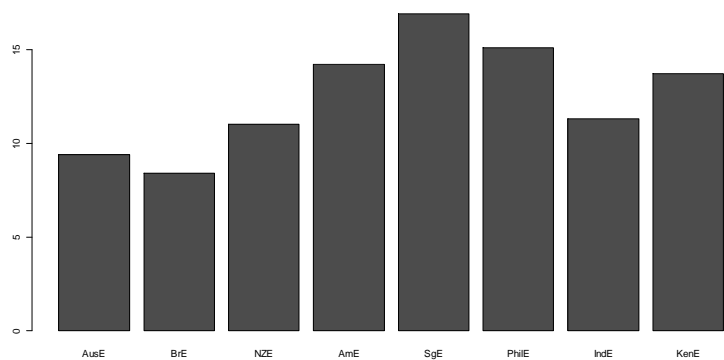
	AusE	BrE	NZE	AmE	SgE	PhilE	IndE	KenE	Total
Textual	9.4	8.4	11	14.2	16.9	15.1	11.3	13.7	100
Structural	9.9	14	16.6	12	9.7	14.1	12.3	11.3	100
Situational	14.7	11.1	14.5	13.1	10.5	15.8	9.6	10.7	100
Factive	12.6	10.1	7.3	28.1	6.7	16.9	6.2	12.1	100
Idiomatic	5.7	7.1	12.1	5	12.1	11.4	9.3	37.1	100
Generic	10	9.2	18.7	8.4	10.8	17.1	5.6	20.3	100
A/Anaphoric	9	9.6	9.6	18.8	16.2	14.3	14.3	8.3	100

It was found that the differences in the number of tokens found in the varieties were significant in all categories.⁵² This means that some varieties make use of a higher or lower number of definites in certain categories than others. Figure 7.3 shows that, however, there emerge no clear patterns of higher or lower use in any of the categories across the nativeness spectrum, a finding that is reminiscent of the patterns seen in chapter 5. Of all the categories, only textual, generic and associative anaphoric show a

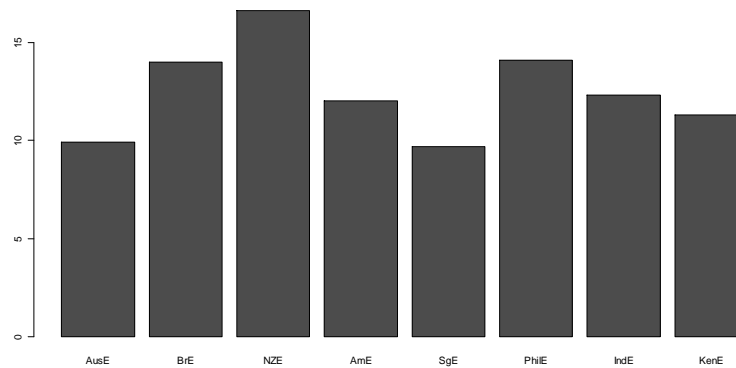
⁵² The chi-squared values for the categories are as follows: textual ($\chi^2(7) = 129.9852$, $p < 0.05$), structural ($\chi^2(7) = 145.6965$, $p < 0.05$), situational ($\chi^2(7) = 161.8779$, $p < 0.05$), factive ($\chi^2(7) = 104.809$, $p < 0.05$), idiomatic ($\chi^2(7) = 84$, $p < 0.05$), generic ($\chi^2(7) = 41.5578$, $p < 0.05$), associative anaphoric ($\chi^2(7) = 45.2932$, $p < 0.05$).

limited degree of tendency to be more frequently used by Outer Circle varieties. Even so, KenE has to be exempted from the overall Outer Circle pattern observed for associative anaphoric.

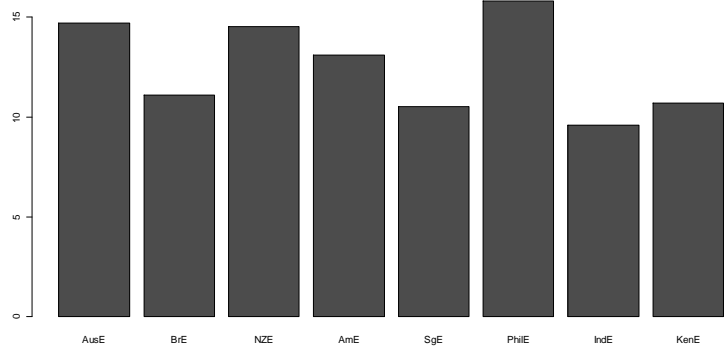
Figure 7.3: A series of bar charts showing the percentages of each category of *the* according to varieties



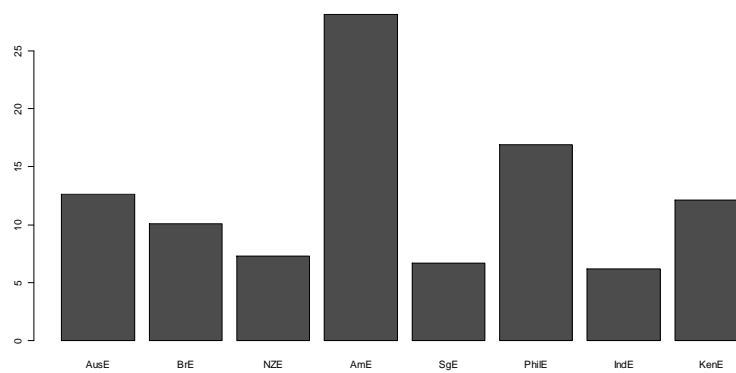
Textual



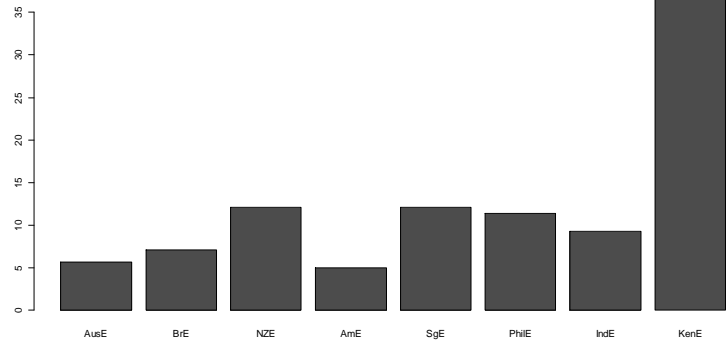
Structural



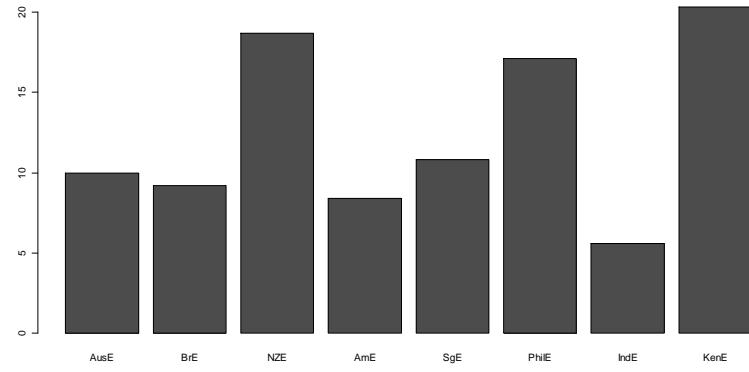
Situational



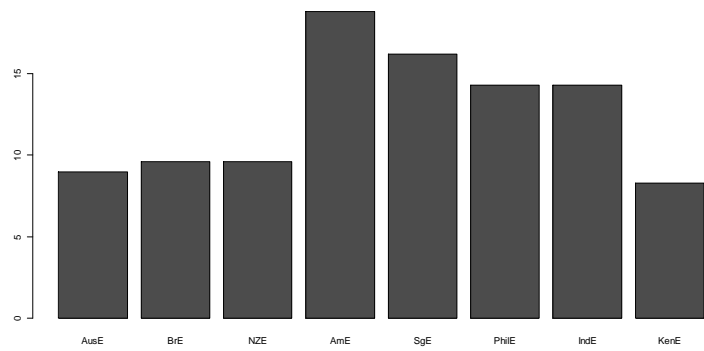
Factive



Idiomatic



Generic



Associative Anaphoric

Based on the patterns obtained in the textual and associative anaphoric categories, it can be said that the text is an important determinant of definiteness for the majority of the Outer Circle varieties. Why this is also a characteristic of AmE is difficult to ascertain. As was done in the previous chapters, however, we can tentatively attribute it to its dissimilar data.

The use of *the* to indicate genericity is also more frequent in the majority of the Outer Circle varieties. Whether or not this is a collective preference is hard to say as it is also found to be a trait of NZE.

As regards the Inner Circle varieties, there is a slight tendency for them to use more situational definites than do the Outer Circle group. However PhilE emerges to be the variety which uses the highest number of tokens in this category.

In chapter five we postulated that article usage types would clearly differentiate the patterns of use in Inner Circle varieties from Outer Circle ones. This was based on the alternative explanations that were offered for Sand's claims (2004) (see 6.1.1 in the last chapter). However, this is not supported by the overall evidence presented in figure 7.3. What the bar charts show us is that the varieties continue to defy any identifiable patterns that we can otherwise use as an indicator of variation. What causes this lack of predictability in their use of the definite article and how do we support the claim made in chapter five that the lexical semantics of nouns is not a reliable predictor of its

variation? We will try to answer the first part of the question below and the second part in the next section.

A claim made by Biber et al. is that variation between genres or text types is more prominent than that between varieties (1999: 20). If the same can be proven through our data, then we have a good explanation for the lack of clear, observable patterns obtained in the varieties.

Based on the data presented in table 7.2, it was hypothesised that the usage types of the definite article and the three categories private conversation (or private dialogue), academic writing and reportage were not independent of each other. The chi-squared tests proved that this was indeed the case and the hypothesis was borne out.⁵³ This means that the text types have each influenced the type of the definite article being used. From table 7.2 it can be seen that the highest or second highest percentages of tokens in the textual, structural and situational categories were from reportage. Generics and associative anaphoric definites appear to favour private conversation in the Inner Circle group but not in the Outer Circle group. Academic writing was more variably used across the usage types.

Speculating about the underlying factors for the varieties' preferences is rather hazardous due to their broad divergence, and such an exercise obviously needs more

⁵³ The chi-square values are as follows: AusE ($\chi^2(6) = 162.1848$, $p < 0.05$), BrE ($\chi^2(6) = 27.0995$, $p < 0.05$), NZE ($\chi^2(6) = 178.8877$, $p < 0.05$), AmE ($\chi^2(6) = 305.6346$, $p < 0.05$), SgE ($\chi^2(6) = 90.8943$, $p < 0.05$), PhilE ($\chi^2(6) = 151.7849$, $p < 0.05$), IndE ($\chi^2(6) = 108.8691$, $p < 0.05$) and KenE ($\chi^2(6) = 88.27$, $p < 0.05$).

data than those analysed for this investigation. Nevertheless it can be said that the aggregate of these differential preferences for usage types by different text types in each variety has a role to play in the varieties' lack of homogeneity. In the following sub-sections we will discuss the qualitative findings of this investigation.

7.3.1 Textual and associative anaphoric *the*

There are no untypical textual and associative anaphoric uses of the definite article found in either in Inner or Outer Circle varieties.

7.3.2 Structural *the*

In some of the Outer Circle varieties, there are a few untypical structural definites which are not found in the Inner Circle group.

In (178), the speaker's use of the PP *of use* which itself contains another PP, i.e. *of languages*, results in *terms* becoming definite. Typical usage would have 'terms' occurring with a zero as in 'in Ø terms of'. What would also be more typical is that the PP *of use of languages* is expressed by the syntactically simpler 'language use' resulting in 'in terms of language use'.

178. A a distinct achievement of Premchand as a craftsmanship regarding his novels being in *the terms of use of languages* of uhm, do you think like that, yes <ICE-IND:S1A-006#99>.

The terms of use is not found in other ICE corpora but occurs twice each in COCA and BYU-BNC. (179) is an example taken from COCA. It is clear that the phrase here refers to an agreement whereas in the speaker in(178) uses it to mean ‘in relation to’.

179. They’ll offer moderated chat in New Zealand and Brazil and will also require United States users to provide them with a credit card number, and this will enable them to track down, hopefully, a user that violates the terms of use
<COCA: 2003/SPOK/NPR_Morning>.

In (180), the use of *the* is categorised as structural due to its occurrence with *appearance (sic) on the stage*. This is a rather ambiguous use of the definite article because the context of this utterance suggests that the phrase could have also been realised as ‘on their very appearance’. Should this be the case, *the* is simply a mistakenly transcribed word and must be discarded. However the postmodifying PP *on the stage* suggests that it is not so and *appearance* must be treated as structural.⁵⁴

180. On the very appearance (sic) on the stage they made other people to laugh where very actions [.] Very small changes on the [...] yes they made jokes
<ICE- IND:S1A-001#206 – 207>.

In (181) the speaker’s use of *the machine CD-ROM* is similar to Hawkins’ *the name Algernon* and thus categorised as structural. Typically this would be realised as ‘the CD-ROM machine’.

⁵⁴ *On the stage* is typically realised as ‘on stage’. This is discussed in 7.3.3.

181. Ya it's those computer is all the data in the computer readable format so you go to the machine CD-Rom <ICE-SIN:S1A-001#188:1:B>.

In the absence of distinct grammatical environments that could be said to influence the untypical uses of structural *the* listed above, coupled with the lack of similar occurrences elsewhere, one is prompted to look at other possible variables in the broader contexts of these NPs. Upon such scrutiny, it has to be said that the speakers' levels of proficiency may have a role to play in the production of these NPs, with (180) being the clearest example. This claim could of course challenge the notion of nativisation that is being investigated. However given the limited spread of structural NPs of this kind, the claim is a highly plausible one and will definitely influence the kind of conclusion that we will draw later. In addition, contrast (179) – (181) with (182), an AmE extract showing an Inner Circle variety speaker who adds the required postmodification through the use of compensatory discourse strategies, thus avoiding further inadequacy in her utterance.

182. LYNNE: and then , so we had to know these tendons, and ligaments, and then, all the bones, ... and, ... you know, all the way up - from the shou=lder down, and then from the hip down. You know, all the way down to the hoof, and then, the bo=ne in the hoof, and, then all the ... wall, and, you know, of the hoof, and, what everything was called there, and then, he went over on how= to trim it [.]

<SBC 131.96 – SBC 151.53>

In the above, notice that structurally definite *the wall of the hoof* used by Lynne is not continuous. First of all, there is a slight hesitation between *the* and *wall*, and as she is about continue, sensing that her listener may not be able to identify the referent of this NP, she decides to add more information (perhaps signalled by the filler *you know*) in the form of a PP, making this a structural definite.

An untypical use of a structural definite that may have little to do with lack of proficiency is shown in (183).

183. B: What is he doing uh Wilmer
C: Uhm Systems Analyst
B: Whoa working where
C: At uhm oh dear this always gives me problems. I think it's called SCO
<laughs> Singapore something something
B: Oh oh okay. I like that Singapore something something. Very informative
C: It's something to do with like uh uh employed by the by Mindef [,] Mindef
B: Oh
D: Oh we went to *the disco there*.

<ICE-SIN:s1a-002#10 – 002#20>

A more typical construction would be 'we went to a disco there'. In (183), however, the speaker chooses to express *disco* as definite by virtue of the postmodifier (*that was*) *there*. It seems that s/he wishes to present this discotheque as possibly the only one in the area being discussed and thus known to his interlocutors. This stands in contrast to the more typical practice of presenting such an entity as a first-mention . None of the

other ICE corpora (or ICE-Sin) has a similar usage but an inspection of COCA and BYU-BNC, yielded four such tokens were found in the former and none in the latter. Below is an example.

184. We all piled into the truck, grateful for the chance simply to sit, and went together to Willmar, and, after we had sold the day's harvest, we went to the lake there and swam away the late afternoon and had a picnic in the shade and drove home at dusk, singing songs or falling happily asleep on our parents' shoulders <COCA: 1990/FIC/Sierra>.

There remains, however, a crucial difference between (183) and (184), as well as the remaining three tokens in COCA. Whereas the *lake* can be easily imagined as part of the countryside setting in the fiction, the *disco* is far from a typical constituent of a ministry of defence (i.e. Mindef). This mismatch of expectations has certainly contributed to the oddness of (183).

In contrast to the above excerpts which were taken from private conversations, the final untypical structural use comes from academic writing. In (185) the use of *the examples of this kind* is found to clearly contrast with the more common 'Ø examples of this kind'.

185. According to Ayer, the examples of this kind are only meant to show how certain class of expressions can be explained in our ordinary language <ICE-IND:W2A-001#87>.

While it may be considered untypical by many, the use of *the* in this case is shown here to be motivated by an otherwise valid principle.

7.3.3 Situational *the*

The situational category produced the highest number of untypical definites in the Outer Circle categories. This is a noteworthy finding given that the quantitative analysis proved that there was no significant difference in the use of this kind of definites between the two groups of varieties.

Excerpt (185) was originally listed above as (181). In this instance we concentrate on *the computer readable format*, which would more typically be realised as ‘a computer readable format’.

185. Ya it’s those computer is all the data in *the computer readable format* so you go to the machine CD-Rom <ICE-SIN:S1A-001#188>.

The speaker chooses to express this NP as definite although the majority of English speakers would present it as a first-mention. This is also the case in (186) – (194). In each of them, there is no prior mention of the entity in the NP in the broader context of the utterance. The only interpretation that can be made for these NPs is that they are somehow related to either the immediate or larger context of utterance.

186. The skin is burning like whereas in Goa we have *the cool climate* <ICE-IND:S1A-001#20>.

187. Naturally British united India for the sake of *the smooth administration* <ICE-IND:S1A-005#49>.

188. Recently, Anil Parab, the gangster who had used *the machine gun* during the killing of Asok Joshi and his goons, near Panvel, got married in the central part of Bombay <ICE-IND:W2C-004>.

189. No something like uhm they used the ones you usually read in Marvel Comics you see the Incredible Hulk fighting the Spiderman <ICE-PHI:S1A-001#230>.
190. we'll just show them how to use a catalogue show them all our library bring them to the different department tell them what to do and all that <ICE-SIN:S1A-001#212:>.
191. A: So last few days
C: Ya, then I went for the designer sale
A: Whoa
A: <laughter>
- <ICE-SIN:S1A-003#X338 – #X341>
192. Rectangle all over the place all over the belt all along the belt and the one thing about it was this gold crest would fall off and I had to go and buy the super glue from from the co-op to stick it back you know [.] So were coming off already I didn't I didn't notice it you know because what happen was was Annie Teo was looking Oh brand name then she just tapped at it and it fell off then I looked at her I said Ooh [.] Then she said not my fault not my fault Ya I think it was about ten dollars worth and it got about six crests on it <ICE-SIN:S1A-003#X354>.
193. <\$C> There are threats no [.] There are threats [.] There are threats [.] Some <- />some <- />some of these kids <-_>Some of these kids<- /> are just carried by the maids <ea/>na <ea/>usiseme [.] No it's true [.] In fact can I tell you something that normally happens one for a maid to do such a thing it's like your

sister you know it's someone they have gotten used to that is another point you have to take into account you see it's not <ea/>ati a maid a neighbour there who comes and rapes a kid of this house all right <ICE-K: S1A003K>.

194. A: No have you been there?

C: No never. Just Bangkok. Ya, just remember the traffic jam that's all.

A: Ya, traffic jams.

C: You tried the you tried that the the toot-toot or something, right?

A: Yes, but only tried in Chiangmai because the rest of the places, we've got, were, were shifted by *the coach*.

<ICE-SIN:S1A-003#140- 146>

Recall from 6.1.1 that the NP *the society* is a rather prevalent usage in the Outer Circle varieties. (196) – (198) are shown here as further illustration.

196. The language used by the <indig> zamindar </indig> or the feudal chief is something <,> high something bombastic something <O> one word </O> and something high sounding but the language used by the servant is something quite different and there you come across a kind of representation of *the society* in terms of the variation of the language itself <ICE-IND:S1A-006#97>.

197. They look to the parties to to influence the course and content of public policy and through their selection of the present and future leaders they really make a big impact on *the society* <ICE-PHI:S2A-043#90>.

198. But that's *the society* [.] Everyone is just opportunistic you know <ICE-K: S1A007K>.

There is also one token of situational definite that more typically occurs with the zero article. The use of *the* in this instance, however, renders *stage* as specific and identifiable and comes across as somewhat odd given the comment being made about the general performance of the persons concerned.

199. On the very appearance (sic) on the stage they made other people to laugh where very actions [...] Very small changes on the [...] yes they made jokes <ICE- IND:S1A-001#206 – 207>.

7.3.3.1 The situational definites and the cline of acceptability

Most of the situational definites listed above have one thing in common: they are not immediately identifiable to the listener. However, it can be seen that some of them are more identifiable than others. *The coach* in (196) is more identifiable than *the machine gun* in (188) which, in turn, is more identifiable than *the cool climate* in (186). To put this observation in context, consider Chesterman's (1991: 65) comment that: "[a]ny use of *the* can be inappropriate – it just depends whom you are talking to, precisely because the appropriateness or otherwise of *the* is defined with relation to the hearer, not the speaker alone. There is, then, a cline of acceptability regarding all definite reference".

When a referent is identifiable to a hearer, then it is 'acceptable' to him/her. If Chesterman's *cline of acceptability* is applied in explaining the untypical definite NPs found in this study, then we will indeed find that, some of them are positioned more closely towards the 'acceptable' end than others. These include, among others, *the coach*, *the designer sale*, *the super glue*, *the machine gun*, *the smooth administration*,

the maids, the different department and the computer readable format. Furthest away would be *the cool climate*. The cline of acceptability is illustrated in figure 7.4.

For the majority of English speakers, upon the utterance of (186) they would sense something amiss in the way that *cool climate* is construed as identifiable since (1) there is more than one type of climate and (2) the cool climate has not been introduced before, an important consideration when one expresses for the first time that s/he possesses such an entity. It would be hard to identify the referent of this NP.⁵⁵

The cline of acceptability will vary from hearer to hearer and this is where the variation really lies. Having said this, let us devote some attention to the more ‘extreme’ untypical definite NPs such as *the cool climate*.

A second language acquisition study by Ionin *et al.* (2004) looks into the acquisition of articles by adult speakers of Russian and Korean and discovers that learners can and do use the definite article as a marker of *specificity* akin to the definite/specific article in Samoan (cf. Lyons, 1999). Ionin *et al.* take specificity to be *speaker intent to refer* (2004: 50). Below are two examples from their data (2004: 48-49):

200. When I was a boy, I found a mine (I mean, an armour, from the World War Two). I liked this kind of things, so I kept it initially in *the secret place* in our yard and then at home.

⁵⁵ However, this is missing Chesterman’s point that all uses of *the* can be inappropriate depending on the hearer. If, in the relevant context, there are only two types of climate being talked about – warm and cool – *the cool climate* makes perfect sense.

202. On Thanksgiving weekend we went to NY for the first time. We took *the room* in the New Yorker Hotel and went outside to see the town.

A similar finding can be found in Goto-Butler's study (2002) of Japanese learners' metalinguistic knowledge in article use. Many of her participants reported that it was consideration of a specific referent that prompted them to use *the*. Ionin et al. attribute the phenomenon to universal grammar and parameter setting (i.e. learners have equal access to [\pm definite] and [\pm specific] but will set their acquisition parameters differently based on one or the other). Based on similar findings in this study, it is fair to say that specificity has influenced the production of situational definites investigated.

It seems that when an untypical use of a situational definite article occurs, the speaker indicates to the hearer that s/he possesses a knowledge of the referent but does not exactly identify it. A different kind of evidence for this claim is found in the following in which the speaker presents an NP (i.e. *fish*) as definite (and specific) but, upon realising the context is insufficient for its identification, retracts by providing a postmodifier in a relative clause.⁵⁶

203. Uhm and they had *the fish which moveable fish* then they had the pagodas it was a earth pagoda and a uh heavenly pagoda one one towards the earth one towards the sky and uh how to put it now <ICE-SIN:S1A-003#X301>.

However, it should be emphasised that it is always the hearer who ultimately determines what is acceptable or not. It must also be added that the use of the definite

⁵⁶ The copula that should come after 'which' is sometimes optional in SgE (Fong 2004:82).

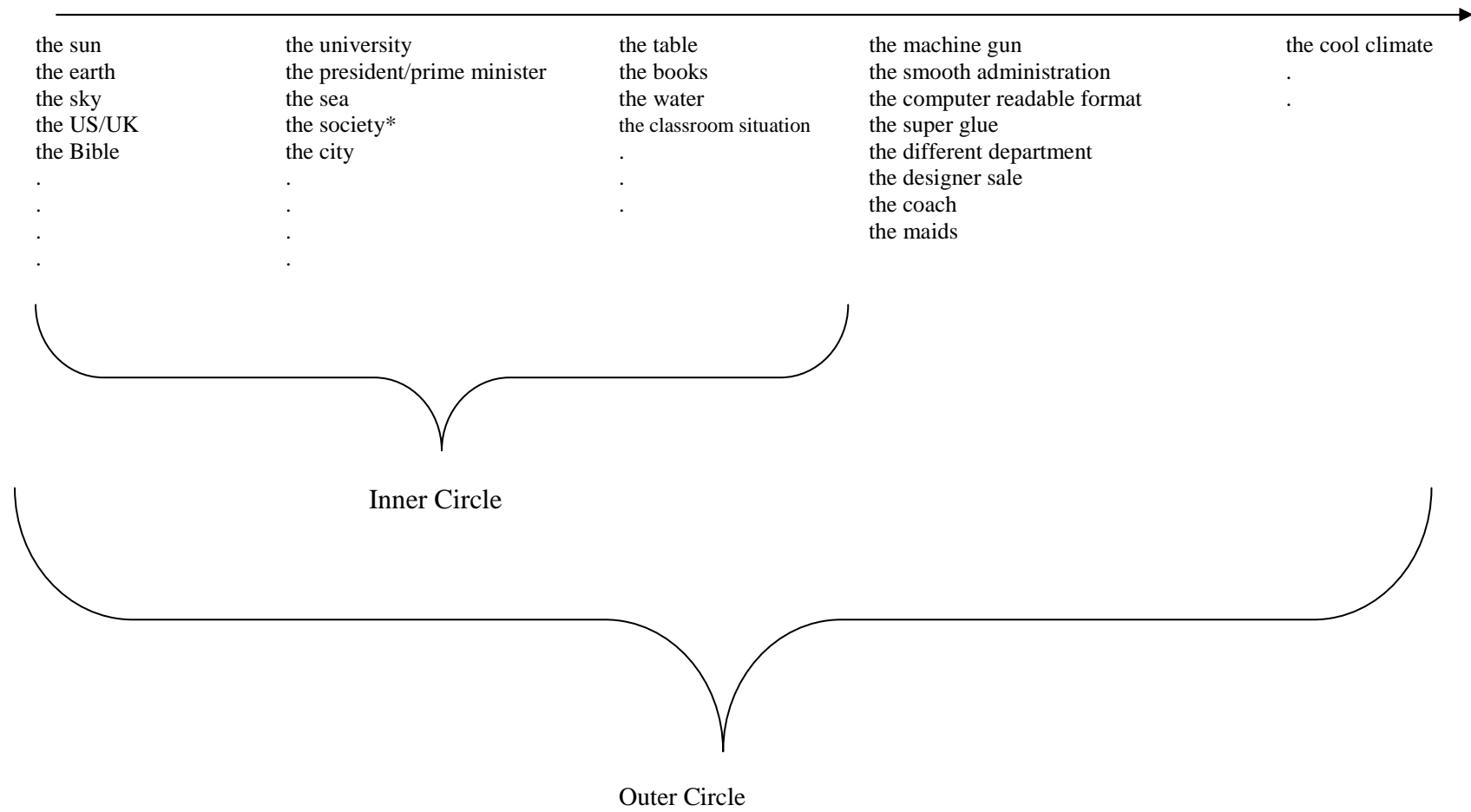
article as a complete specificity marker is limited to the more extreme cases of untypical uses.

Figure 7.4 shows that the most identifiable/acceptable situational definites are Hawkins' universal ones e.g. *the sun*, *the earth* and *the US*, followed by entities known through specific knowledge in a larger situation, e.g. *the university*, *the president* and *the sea*. Then these are followed by entities identified by the immediate situation, e.g. *the books*, *the table* and *the classroom situation* (uttered in a school). As the cline becomes progressively more unacceptable, so too does the identifiability of the NPs that are found along it. Based on the data, it seems that the referents of these less identifiable NPs are associated with the larger situation by the speakers. The burden of their identifiability, however, is shifted to the listeners who now appear to bear the task of working out the connections. It is useful to bear in mind that although the less acceptable NPs are shown to be associated with Outer Circle varieties only, this by no means implies that these varieties are equal in their level of use of such NPs.

Figure 7.4: The cline of acceptability of the situational definite article

Most acceptable/identifiable

Least acceptable/identifiable



Note that the above untypical usages stem from the use of *the* instead of *a/an* or zero. When *the* is used instead of null, there are two possibilities. If the referent remains identifiable due to its situational context, then it is placed in the acceptable end of the cline. This is the case of *the society* (shown in the figure) and *the Spiderman*. The case of *the stage* is rather different. On hearing the utterance in (199), if a listener understands that the comment being made about the performers is general in nature, then *the stage* will turn out to be quite unidentifiable and is placed towards the unacceptable end of the cline.

7.3.4 Factive, generic and idiomatic *the*

No untypical instances of these kinds of definite NPs were found in the data.

7.4 Results: *a/an*

A total of 5105 tokens of *a/an* were analysed in all corpora. 2953 tokens were from the Inner Circle group while 2152 were from the Outer Circle group. Table 7.4 shows the breakdown of the analysed tokens according to varieties and categories.

The table shows that *a/an* is used most frequently in the first-mention/quantitative category. This is hardly surprising, given the origin of this determinative from the numeral 'one'. The second place is occupied by factive, which is followed closely by idiomatic. In contrast to the definite article, the indefinite *a/an* is more suitable for use in the factive category in which specificity is not crucial. *A/an* is also frequently used in idioms due to its appearance in many non-numerical quantifiers. The lower occurrence

of this article in the generic category as compared to the definite article is somewhat unexpected, however. This is based on claims that the definite article is not extensively used outside the scientific register (Parrish, 1987; Tarone and Parrish, 1988 and Whitman, 1974 in Liu and Gleason, 2002: 5-6). Based on the data, it seems that *the* is more favoured than *a/an* in expressing genericity.

As was the case for the definite article in 7.3, in order to find out if the varieties have used significantly more or fewer tokens in certain categories than others, it is necessary to determine the extent to which each category is constrained by the varieties. Table 7.4 shows the comparison in this perspective with figure 7.5 illustrating the differences with a corresponding series of bar charts.

Table 7.4: Tokens of *a/an* according to varieties and categories

		Idiomatic	FM/Quant.	Factive	Generic	Total
AusE	Priv. Convs.	35 (18%)	122 (63%)	36 (19%)	0 (0%)	193 (100%)
	Acad. Wrtg.	15 (10%)	121 (80%)	16 (11%)	0 (0%)	152 (100%)
	Reportage	10 (5%)	175 (85%)	21 (10%)	11 (5%)	207 (100%)
	Total	60 (11%)	418 (76%)	73 (13%)	11 (2%)	552 (100%)
BrE	Priv. Convs.	56 (18%)	187 (60%)	68 (22%)	0 (0%)	311 (100%)
	Acad. Wrtg.	14 (7%)	163 (85%)	13 (7%)	2 (1%)	192 (100%)
	Reportage	15 (6%)	185 (79%)	33 (14%)	1 (0.4%)	234 (100%)
	Total	85 (12%)	535 (73%)	114 (15%)	3 (0.4%)	737 (100%)
NZE	Priv. Convs.	33 (15%)	145 (66%)	38 (17%)	3 (1%)	219 (100%)
	Acad. Wrtg.	31 (9%)	274 (80%)	33 (10%)	5 (1%)	343 (100%)
	Reportage	21 (9%)	206 (83%)	18 (7%)	(1%)	247 (100%)
	Total	85 (11%)	625 (77%)	89 (11%)	10 (1%)	809 (100%)
AmE	Priv. Convs.	102 (23%)	201 (46%)	106 (24%)	29 (7%)	438 (100%)
	Acad. Wrtg.	12 (7%)	140 (78%)	24 (13%)	4 (2%)	180 (100%)
	Reportage	4 (2%)	200 (84%)	31 (13%)	2 (1%)	237 (100%)

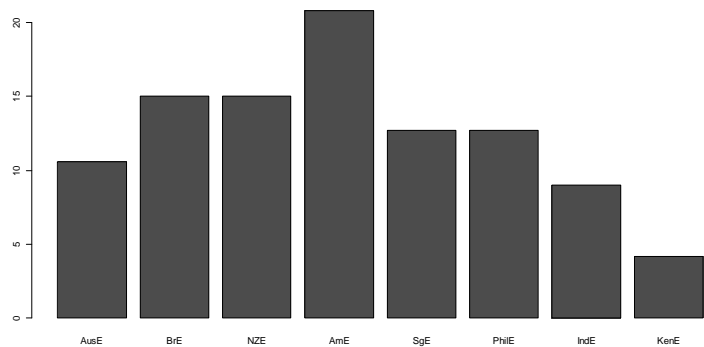
	Total	118 (14%)	541 (63%)	161 (19%)	35 (4%)	855 (100%)
SgE	Priv. Convr.	35 (18%)	106 (56%)	50 (26%)	0 (0%)	191 (100%)
	Acad. Wrtg.	12 (7%)	132 (77%)	20 (12%)	7 (4%)	171 (100%)
	Reportage	25 (13%)	139 (73%)	26 (14%)	1 (0.5%)	191 (100%)
	Total	72 (13%)	377 (68%)	96 (17%)	8 (2%)	553 (100%)
PhilE	Priv. Convr.	35 (16%)	120 (54%)	67 (30%)	0 (0%)	222 (100%)
	Acad. Wrtg.	34 (15%)	175 (75%)	25 (11%)	1 (0.4%)	235 (100%)
	Reportage	3 (2%)	153 (87%)	19 (11%)	1 (0.6%)	176 (100%)
	Total	72 (11%)	448 (71%)	111 (18%)	2 (0.3%)	633 (100%)
IndE	Priv. Convr.	16 (8%)	130 (62%)	65 (31%)	0 (0%)	211 (100%)
	Acad. Wrtg.	27 (15%)	103 (57%)	44 (24%)	6 (3%)	180 (100%)
	Reportage	8 (6%)	114 (86%)	9 (7%)	1 (0.8%)	132 (100%)
	Total	51 (10%)	347 (66%)	118 (23%)	7 (1%)	523 (100%)
KenE	Priv. Convr.	7 (5%)	94 (67%)	39 (28%)	0 (0%)	140 (100%)
	Acad. Wrtg.	11 (7%)	126 (77%)	26 (16%)	1 (0.6%)	164 (100%)
	Reportage	6 (4%)	114 (82%)	19 (14%)	0 (0%)	139 (100%)
	Total	24 (5%)	334 (73%)	84 (19%)	1 (0.2%)	443 (100%)

Total	567 (11%)	3625 (71%)	846 (17%)	77 (2%)	5105 (100%)
-------	-----------	------------	-----------	---------	-------------

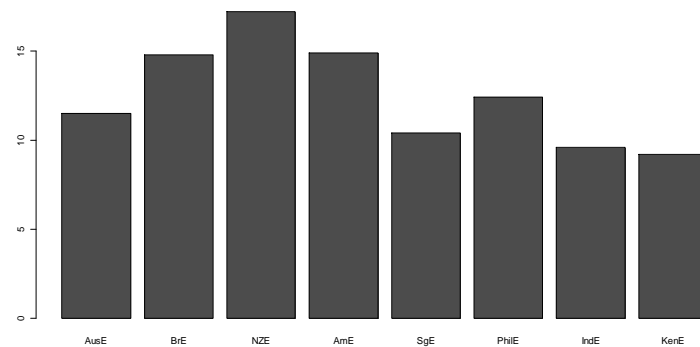
Table 7.5: Percentages of each category of *a/an* according to varieties

	AusE	BrE	NZE	AmE	SgE	PhilE	IndE	KenE	Total
Idiomatic	10.6	15	15	20.8	12.7	12.7	9	4.2	100
FM/Quant.	11.5	14.8	17.2	14.9	10.4	12.4	9.6	9.2	100
Factive	8.6	13.4	10.5	19	11.3	13.1	13.9	10	100
Generic	14.5	3.9	13	44.5	10.4	2.6	9.1	1.3	100

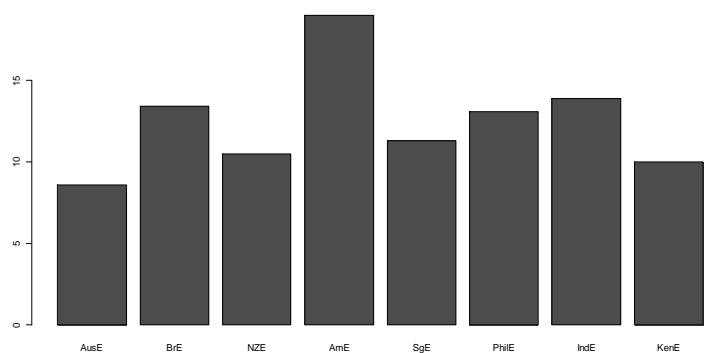
Figure 7.5: A series of bar charts showing the percentages of each category of *a/an* according to varieties



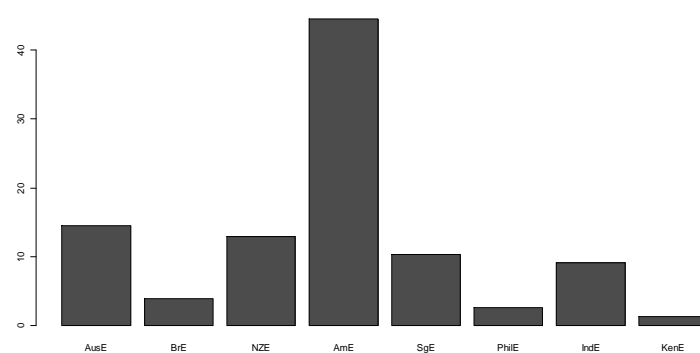
idiomatic



first-mention/quantitative



factive



generic

Figure 7.5 again shows that the patterns of usage types of *a/an* cannot be neatly divided between Inner and Outer Circle varieties. The chi-squared statistic shows that the differences between the varieties are significant.⁵⁷ There are some weak patterns observed for idiomatic and first-mention/quantitative i.e. the Inner Circle group tended to make more use of these categories than did the Outer Circle. There is a greater variety and a higher use of non-numerical quantifiers found in the Inner Circle group than in its Outer Circle counterpart and this may have influenced the overall higher percentage in the former (see table 6.8 in chapter 6).

As for the first-mention/quantitative category, it is harder to speculate about its causes. As we saw that there is no corresponding increase in the use of *the* in the situational category in 7.3, it would not be sensible to suggest that first-mention/quantitative *a/an* has been systematically replaced by the definite article. However, we do not reject the possibility that other determinatives e.g. 'this', 'that', 'one', etc. may take the place of *a/an*.⁵⁸ Nevertheless, its confirmation lies beyond the scope of this thesis.

AmE continues to defy the patterns of usage of other Inner Circle varieties by surpassing their percentages in three of the four categories. Given the previous findings of this and earlier chapters, we will treat its demonstrated trend with caution by tying it to the overall trend shown by the three other Inner Circle varieties.

⁵⁷ The chi-squared values are as follows: idiomatic ($\chi^2(7) = 75.2434$, $p < 0.05$), first-mention/quantitative ($\chi^2(7) = 168.7721$, $p < 0.05$), factive ($\chi^2(7) = 49.357$, $p < 0.05$), generic ($\chi^2(7) = 86.4286$, $p < 0.05$ with Yates' correction).

⁵⁸ But this does not include the zero article as this possibility was ruled out based on the analysis of overall frequencies in chapter five.

Next we examine the percentages of the usage types in relation to text types. However, the generic usage type is not included in this analysis because three of its cells have a value of 0 and five other cells have a value of less than five, making it highly unsuitable for the test of independence. The chi-square test results show that differences in the percentages are significant. This means that text types are related to the use of usage types of *a/an* (and possibly, vice-versa). Private conversation is seen as predisposed to factive usage while reportage tends to increase the use of first-mention/quantitative *a/an*. Academic writing is shown to have a weak predilection for idiomatic usage although in a few varieties this place is taken by reportage. Once again, Biber's et al.'s (1999: 20) claim regarding the superiority of text types over varieties in influencing variation is substantiated. What, then, can be said about the role of varieties?

The next section presents some qualitative findings of the usage types of *a/an* to examine in detail if the varieties do give rise to variation.

7.4.1 Idiomatic *a/an*

As it was for the definite article, there are no untypical tokens of idiomatic *a/an*. This is probably because of the 'frozen' nature of idioms which allows no variation.

7.4.2 First-mention/quantitative *a/an*

There are a few untypical tokens of first-mention/quantitative *a/an* found in the Outer Circle varieties. The first type comprises tokens of *a/an* that are more typically realised as *the*, as in:

204. No self-control [.] We have [...] But I told you I put on that mini by the way they even go to an extent of tearing it you know that means you can't control yourself because what you are seeing is not you want to see something more than that [...] If your eyes [...] Yeah Yeah <ICE-KEN: S1A003K>.
205. There is [,] there is a uh club [,] this humour club [,] oh good [,]and <w> I've </w> chance I had a chance of uh [,] attending this humour club <ICE-IND:S1A-001#193>.

It can be seen in (203) and (204) that the choice of *a/an* over the more typical *the* has changed the meanings of the NPs concerned. To understand the semantic change that has taken place, we have to draw on Chesterman's theory of definiteness (1991/2005) (see 2.7.2). Recall that definiteness is viewed as a composite of three meanings: locatability (familiarity), inclusiveness (quantity) and extensivity (generality).

In (204) and (205), the quantitative meaning of *a/an* expressed is not foremost. Rather it comes across as a first-mention in which identifiability is not established. While typically we would expect the NPs to be realised as *the extent of tearing it* and *the chance of attending*, we now see that the speakers have chosen not to construe *extent* and *chance* as inclusive, in effect implying that there are other *extents* and *chances*.

The same mechanism is applied by the speaker of (206), although this time it involves a nominal that more readily presents itself as one of the many, i.e. time. This usage is absent elsewhere in the ICE corpora investigated but two similar tokens were found in COCA and six tokens were found in BYU-BNC.

206. So [,] then then he could read he read a lot then he could write he could think it was like you know when you're a very busy person and suddenly you're given a time to do what you always say oh yeah if I had the time I could do it <ICE-PHI:S1A-005#179>.

(207) is an example from BYU-BNC. Despite the similarity between the underlined NPs in (206) and (207), notice that the Inner Circle usage clearly implies that *a time* refers to a time slot. On the other hand, the same NP is used to refer to an opportunity in the Outer Circle usage.

207. And you was given a time to go in to sign on. We had to sign on er three times a week. We was er given a time to go and er it was anything after nine o'clock, in the morning. I see. And then you got You had to go three times a week? Three times a week, then, yes. Why was that, then? Well, it was starvation years, actually <BYU-BNC/FY2/S_interview_oral_history>.

Untypical uses of *a/an* in the Outer Circle varieties can also imply that a certain recategorisation of countness of a noun has taken place. This is illustrated by *a new staff* in (208).

208. No I don't [.] For those teaching staff they as and when they come in a new staff we'll do <ICE-SIN:S1A-001#209>.

Notice that the usually mass noun of staff is rendered singular by the indefinite article, thus signalling 'a new staff member'. More untypical uses of this kind are shown in (210) – (212).

210. Sometimes there is a police <ICE-IND:S1A-001#48>.

211. So this type of lively picturise then gives you a very healthy and lively informatives he used to give [.] yes [.] regarding his location <ICE-IND:S1A-006#75>.

212. Apparently they can tape your phone conversation and use it as a what evidence in court ah <ICE-SIN:S1A-005#285>.

The NP a police in (210) is likely to refer to a policeman but may imply a police force. In (211) 'information' is presented as *informatives* resulting in what can be interpreted as a single piece of information.⁵⁹ *A what evidence* implies a piece of a certain kind of evidence in (212). Not all such uses of *a/an* force a rather radical recategorisation of countness of the nouns involved. In (213) *a sauce* is common in Inner Circle varieties. The difference is a sauce in these varieties is usually used to denote 'a kind of sauce' while *a tomato sauce* in (213) most likely means 'an amount of (tomato) sauce', it already being a specific kind of sauce.

⁵⁹ It is not known if /s/ is a morpheme for the speaker or simply a mispronunciation of *informative*.

213. Sometimes I just [,] just fry a bit slightly and then [,] put you know [,] I just put into *a tomato sauce* and onion then ginger and garlic and big chillies or capsicum whatever <w> you've <ICE-IND:S1A-007#203>.

Recategorisation is a quite productive feature even in Inner Circle varieties in which we can sometimes find instances of *researches*, *luggages*, *staffs*, and so on. However, notice that each of these is a case of pluralisation. What makes the untypical use of *a/an* in Outer Circle varieties noteworthy is the tendency to perceive these otherwise noncount or mass entities as a single instance of a kind, resulting in, for example, *a staff* for a staff member or *a police* for a police officer. The corpus data do not allow us to speculate about the causes of this tendency but they may have something to do with the way the same referents are construed in the speakers' native languages. In Malay, for example, an equivalent to 'staff' would be 'worker', which is singular. The word 'staff' itself has been included in the Malay lexicon to denote either 'worker' or 'workers'. When *staff* is used by such a speaker, the singular concept of the noun in Malay may persist into English.

However, mention must also be made of the fact that, in the data, almost all untypical uses of first-mention/quantitative *a/an* that recategorise countness have their referents premodified as in *new staff*, *healthy and lively informatives*, *what evidence* and *tomato sauce*. This premodification appears to strengthen the need to present the otherwise noncount entity as an instance of a kind. This, however, is also a characteristic of Inner

Circle varieties. Quirk et al., for example, note that ‘a sensitivity’ is less acceptable than ‘a charming sensitivity’ (1985: 287).

7.4.3 Factive and generic *a/an*

These two categories did not yield any tokens of untypical uses.

7.4.4 Other untypical characteristics of *a/an* in Outer Circle varieties

The following are tokens that exhibit a degree of untypicality that is not central to the meaning of the article itself but are manifested more than once in the data.

214. That means you can’t have *an understand* [,] but [,] you should keep that others you have to solve that [,] problem of with management <ICE-IND:S1A-003#258>.

215. All the happening [,] *a real happening* of the society <ICE-IND:S1A-006#146>.

The above excerpts indicate that *a/an* may at times occur with a head that lacks a noun quality or is vague in regard to its sense. In (214), in what appears to be a slight overall deficiency in grammaticality on the speaker’s part, ‘understanding’ is realised as *understand*, making it identical to the corresponding verb. In (215), *happening* is more likely to refer to ‘story’ or ‘incident’ (notice the previous phrase *all the happening* which provides some of the context) but not a spontaneous activity or an adjective which means being trendy. The use of *a/an* is regarded as unusual because the choice of *happening* quite often refers to the adjective.

Other than the above the choice of *a* or *an*, which is usually dictated by the presence or absence of a vowel onset in the pronunciation of the succeeding word in the NP, is sometimes made without this consideration. This can be seen in (216) and (217).

216. Uhm and they had the fish which moveable fish then they had the pagodas it was a earth pagoda and a uh heavenly pagoda one one towards the earth one towards the sky and uh how to put it now <ICE-SIN:S1A-003#X301>.

217. But fish in itself doesn't contain as much as cholestrol as a egg or meat <.,> <ICE-IND:S1A-007#138:1>.⁶⁰

7.5 Results: Ø

This section discusses the findings of Ø based on a case study of the word 'house'. Table 7.6 presents the breakdown of tokens according to determination types and varieties.

Table 7.6 *House* according to determination types and varieties

	AusE	BrE	NZE	SgE	IndE	PhilE	KenE	Total
Determined	350 99%	407 99%	551 99.6%	295 98%	396 94%	344 98%	571 99.5%	2914 98%
Undetermined	3 1%	6 1%	2 0.4%	6 2%	26 6%	8 2%	3 0.5%	53 2%
Total	353 100%	413 100%	553 100%	301 100%	422 100%	355 100%	574 100%	2967 100%

⁶⁰ Based on my experience with speakers of IndE, however, vowel onsets are occasionally realised with /y/, which is a glide and requires *a*, e.g. 'a yegg' instead of 'an egg'.

It can be seen from the above table that Outer Circle varieties have a slight tendency to use more undetermined NPs headed by 'house'. However this does not necessarily mean that all such uses are untypical. Many of these uses are governed by the principles described in 2.1.3 (see chapter 2). The following examples are taken from the Inner Circle varieties.

218. They are moving Ø house today - into the portion below Greg [and] Ken's flat cum office [.] They have 4 small children (7yrs - 9 mths), and if when Ken goes back to N.Z , s (sic) Steve will share the Greg's portion with him as an office . They'll use Fitches' green van to move the big furniture today <ICE-AUS: W1B- 002(noone):30>.
219. it was hard to see in the house [,] it was a dark sort of Ø house <ICE-NZ:S1B-069#230 – 231>.
220. I cleaned Ø house for a couple in Goathouse street and another lady in Park drive <ICE-AUS:S1A-034(C):176>.
221. B: I was totally infatuated by him,
A: So you set up Ø house together

<ICE-GB:S1B-049 #137-138>
224. With regular television (cable and satellite have different problems), reception (sic) quality can vary from country to country, and even from Ø house to Ø house in the same street <ICE-GB:W2D-014>.
225. 'Ø Last house above the harbour under the mountain, y' can't miss it. It looks like shit!' The phone clicked off in her ear <ICE-AUS: W2F-004(noone):131>.

While it can be argued that some of the tokens above are idiomatic, the undetermined status of ‘house’ still needs to be accounted for. To do this we make use again of the composite meaning of definiteness by Chesterman (1991/2005). In (218) the phrase ‘move(ing) house’ is used to denote ‘move to a different house’. The idiomaticity of the phrase is contained in the fact that the quality of the house as ‘different’ and the motional meaning of ‘to’ are expressed in it despite not being overtly manifest. At the same time, however, the semantic weights of ‘move’ and ‘house’ can be said to be equal and thus rules out the possibility of the phrase being a case of denominalisation or noun incorporation. In view of these facts, the most likely argument that can be made for it is that the use of *house* occurs with a null article (see 2.5.3). This means that *house* in *moving house* is (i) locatable (ii) inclusive and (iii) unlimited in extensivity (see table 2.3). Put differently, the house in this case (i) is not physically bounded (i.e. conceptual) (ii) is an ‘institution’ instead of a discrete item and thus (ii) identifiable.⁶¹ This claim is supported by the fact that *house* would require the definite article if it were to take restrictive modification, i.e.

218a. They are moving \emptyset house today. The house that they’re moving into is the portion below Greg and Ken's flat cum office.

Another way to corroborate the claim that *house* is abstract is to point out the fact that the house in the sentence is not a prototypical house but is referred to as ‘a portion’.

⁶¹ It is for the same reasons that I disagree with Wee and Umberto’s analysis of *watch movie* in ‘OK, you want to go to watch \emptyset movie or you want to ...’ in colloquial SgE as noun incorporation (2004: 58-60). They argue that “the individual salience and referentiality of *movie* as a separate entity from the verb *watch* has been reduced so that the speaker is highlighting the activity of ‘movie-watching’ (ibid.). On our analysis, *watch* and *movie* carry equal semantic weights. *Movie* would be seen to occur with the null article, and would thus be an abstract concept. (218), (220) and (221) are evidence that it occurs in more standard varieties as well. My data do not contain tokens that are similar to theirs, and hence this footnote.

(220) and (221) are similar cases. The Outer Circle varieties, with the exception of IndE, have such uses of ‘house’ as well. These are exemplified by (226) and (227). \emptyset in (219) is also a null article (cf. the use of null in ‘kind-of’ contexts in Chesterman, 1991/2005: 73) but the phrase ‘a sort of \emptyset + singular N’ appears to occur only in Inner Circle varieties with just two similar tokens found in SgE. One is shown in (228).

226. Patriarchal conjugal relations are not questioned either in both plays; the husband goes off to work and the mother stays home to keep \emptyset house, hiding the reality that women participate in productive activities in these rural households <ICE-PHI:W2A-002#74>.

227. We well we got her when she was like six months old and someone was moving \emptyset house and didn't want her <ICE-SIN:S1A-039#22:>.

228. A plump, pleasant-faced mother of two boys, she was a practical sort of \emptyset person who did not believe in procrastination <ICE-SIN:W2B-005#178:1>.

(224) is also a case of *house* occurring with the null article (see ‘parallel structures’ in Chesterman, 1991/2005: 5). However the above analysis is slightly problematic for this case considering two \emptyset *houses* are involved. One way to circumvent the restrictive interpretation of only one abstract concept of ‘house’ is involved in the sentence is to allow two ideas of ‘house’, one qualitatively different from the other (e.g. ‘small house’ idea from ‘big house’ idea), to take up the roles of the two entities that is required by the structure ‘from N to N’. This analysis appears to be supported by (229) found in COCA although in this case two entirely different modes of transport are involved:

229. Holmes was one gritty gal who would go on to star in other popular serials, her specialty being the ability to run along the top of boxcars and to leap from train to horse <COCA:1992/News/SanFrancisco>.

(230) is the only token available from the Outer Circle, although the preposition ‘from’ is only implied from the context:

230. In order to answer the above questions a stepwise series of field studies to examine natural variation at different micro-environmental levels (Ø house to Ø house) will be conducted at selected sites <ICE-KEN: W2A027K>.

Finally, (225) shows the use Ø house that is due to omission as a matter of convenience, not because there is a special meaning is intended by omitting the definite article. Newspaper headlines makes frequent use of this feature (see ‘block language’ in Biber et al., 1999: 263). One such use of Ø house is found in the Outer Circle varieties. In (231), Ø house is used a shortened form of the House of Representatives in the Philippines.

231. Ø House frowns on new taxes <ICE-PHI:W2C-019#54>.

Untypical uses in the Outer Circle varieties are listed below:

232. Actually, we will be more than full Ø house <ICE-PHI:W1B-003#79>.

233. Between October, 1991 and July, 1992 the accused never came to my house. I knew Ø house of the accused at Jericho when we were still friends <ICE-KEN: S1BCE05K>.

234. The other issue that we need to look into is that of Ø condemned house, more so in Nairobi. It has become a habit - and this has been an audit query for the last five years - that some officers, including Permanent Secretaries and even some hon. colleagues here, get houses condemned. <ICE-KEN:S1B057HK>.
235. B: But would it be worse in a in a terrace house or in <unclear> word </unclear>?
- A: But but Ø terrace house uh at least uh when you come out to the garden or you dispose off rubbish at least you can still catch one another you know
- <ICE-SIN:S1A-094#124-125>
236. That was the Three Cables. Where got three sisters. The house is called Three Cables where old woman sits in Ø house <ICE-SIN:S1A-030#X352>.
237. So it means that uh you you find people saving up for Ø house some of them even want a landed property before they get married and uh it may be a bit uh too much of a high expectation but there it goes you know uh people do think that way <ICE-SIN:S1B-025#51>.
238. Had the boy in Ø house not seduced her first time her life would have been different <ICE-IND:W1A-018#96>.
239. They leave their childrens in Ø house. When mother [and] father go to work children were became alone. <ICE-IND:W1A-004#74>.
240. C: All friends
- B: At Ø house only or at hotel

C: At Ø house only [.]

<ICE-IND:S1A-051#206-208>

241. Otherwise all these girls what they will do in college [.] they will do only studies [.] and then they will get marry [.] And afterwards it will be very difficult for them to adjust all this [.] means uh [.] problem so now [.] I'm teaching her all [.] Ø house <ICE-IND:S1A-043 – 043#75>.

242. In Ø house whom when mother is learned, literate then her children [...] [.] are also intelligent <ICE-IND:W1A-004#117>.

243. A: I don't know man [.] he not [.] I'm not meeting him now-a- days [.] not seen him

B: He is in Ø house only no

<ICE-IND:S1A-051#124 – 125>

244. C: But my house is in Tirumangalam

A: Tirumangalam. How far about [...] ?

C: Twenty-two kilometres from Ø house

A: Twenty kilometres

<ICE-IND:S1A-024#56 – 60>

245. When mother is not in Ø house father would have to take care of children <ICE- IND:W1A-004#104>.

246. Women lived only in Ø house <ICE-IND:W1A-004#56>.

247. <indig> Shri </indig> Suvarna with the permission of Ø house <,> read the kannada version of the report <ICE-IND:S1B-078#25>.

248. The distance between Kantilya Sadan and Ø house of that Raj [,] Ø house of Rajan is about five minutes walk on foot <ICE-IND:S1B-068#275:1:B>.
249. And the president for that [,] for the last fifteen years [,] while acting also I was so busy as a artist even then I used to uh give lot of service to the [,] especially this old indigent (sic) artist giving pensions [,] and then getting Ø house <O> one word </O> for them [,] <ICE-IND:S1B-043#68>.
250. E: That's obviously [,] I suspected only one person and visited therefore Ø only house
- C: I suspected
- E: Only one person
- <ICE-IND:S1B-065#172 – 174>
251. <O> One word </O> you will feel bored <indig> jaar </indig> [,] because this not Ø house <indig> jaar </indig> [,] come on [,] <ICE-IND:S1A-054#89>.
252. There [,] I'll be spending my time in Ø house <ICE-IND:S1A-016#31>.

(232) is an adjectival use of the NP 'full house' that is usually used to describe a room or a building which has reached its maximum occupation limit. This meaning is used here to basically mean 'full'.

(233) and (248) are tokens of *house* without the required definite article. Notice that these postmodified NPs contain enough information that could have led the speakers to omit the definite article. This is similar to a proper name treatment. From COCA's spoken data, such tokens are exemplified below.

253. Mrs. SANTOS That's right. It's a fabulous question because I don't need to do anything. I can just sit home and polish my nails and look up at the ceiling. I've always been working. I enjoy achieving. It's not a drive per se. It's just I like to do something with my life, and five years ago I had started with Home Shopping Network and everybody said are you kidding? And now everybody is on the channel. So House of Ivana is growing, it's a company which me and Ricardo... Mr. HAHN House of Ivana? Mrs. SANTOS House of Ivana, uh-huh, which me and Ricardo are partners on and we have a line of clothes and jewelry and cosmetics and accessories, and we have manufacturing facilities all around the world.

<COCA:1997/SPOK/CNN_King>

The use the null article instead of *a/an* in several of the listed excerpts above carries the implication that the referents of the NPs are not singular but mass (see *moving Ø house* above). This means that number distinctions are not marked in the nouns concerned. This result is most clearly seen in (234) where *condemned house* is unambiguously referred to in the plural shortly later in the same discourse. Number marking is less crucial in (235) and (237); however, they still give out the meaning that *house* is a concept, rather than a bounded entity. Of interest is a token in NZE which displays the same meaning that is discussed here:

254. Further, the quality of the house tends to be related to rank: the higher the rank the better the quality of Ø house <ICE-NZ:W2A-017#79>.

There are no other similar tokens of this type in the other Inner Circle varieties (or in COCA and BYU-BNC).

(244), (249) and (251) are also cases of abstract *house* similar to (235) and (237). The significance of Chesterman's concept of extensivity is especially evident in (244). In it, *house* in *twenty-two kilometres from house* is construed as the only concept of house that exists in the discourse universe of the interlocutors. If this is not assumed, the information regarding the distance to Tirumangalam will not be useful to the listener.

Many of the above excerpts involve the phrase 'in house'. This is an interesting characteristic considering that 'in + house' is a widely used compound which describes an activity being carried out in an establishment without external agency or participation. The 'house' in the compound is a noun but has an equal semantic weight with the preposition. The outcome of this combination is an adjective. This is the pattern found in the Inner Circle varieties and the majority of texts in their Outer Circle counterparts, as can be seen in (255):

255. Areas where policies and procedures may be needed include: <indent>network resilience (how the service will be supported in the event of network failure) [and] semi; network provision, for example, whether the network should be provided by: <indent>purchasing network services from a vendor or using *an in-house network* [...] <ICE-NZ:W2D-010#25>.

However, some speakers of the Outer Circle varieties have extended the meaning of this adjectival compound to include residential houses, as opposed to companies or

professional organisations. This gives the impression that *in house* as used by these speakers is the use of ‘in + Ø house’ as a PP when, in fact, the phrase is an adjective. It appears to be the case in (243) and (245) where the copula is present. Nevertheless, true to its adjectival nature, *in house* can be transformed into an adverb, albeit without the usual derivative marking, as demonstrated in (236), (239), (246) and (252).

Despite this analysis, we still have to account for *in house* in (238) and (242). Looking at the contexts of the sentences, it appears that the tokens are PPs, with the article (*the* in 238 and *a* in 242) omitted. What the null article has done to the head in each case is to render it boundless, thus making it maximally extensive or abstract. Again the concept of ‘house’ is the foremost meaning and identifiability of the referent is possible in both cases. It must be pointed out *house* in (242) is also impervious to number. Therefore an alternative interpretation of this sentence would be ‘women lived only in houses’. This would make zero as the article in use. Should we choose to interpret (242) this way (i.e. and the NP is thus not marked for plurality), zero is seen as replaced by null, resulting in *house* being identifiable. (238) is also a case of *house* with the null article in a PP. However it contains two such tokens.

(247) differs from (231) in that it is not an instance of block language. *House* in this sentence refers to one of the two houses in the Indian parliament. Thus typically it requires the definite article. In its place, however, the null article is used. As in earlier cases, this too lends identifiability, inclusiveness and maximum extensivity to the

referent. However given the context of the discourse in which the house is a one-member set, it is not turned into an abstract of idea of 'house', but a proper name.

Finally, (241) and (250) can also be submitted to the same analysis described here. However, taking into consideration the overall grammatical quality of the utterances, we need to be careful about postulating any principles that can be said to underlie the uses since they can, in fact, be no more than genuine errors.

7.6 Conclusions

The study of articles described in this chapter was based on the corpus-based technique of annotation. It was intended to investigate if the use to which an article is put had any effects on its patterns of use in the varieties. To recapitulate, an annotation scheme for the definite article was developed based on the categories presented in Hawkins (1978) and Prince (1981; 1992). Another annotation scheme was developed for *a/an* using information available in major grammars. For the study of \emptyset , a case study of 'house' was conducted.

Regarding the definite article, there were significant differences in the percentages of usage types across the varieties. However there was no clear trend that could be observed in both the Inner Circle and Outer Circle groups. It was found that text types or genres influenced the usage types more than did the varieties. Reportage had a preference for structural, situational and textual usage types more than did private conversation and academic writing. Because little could be said about the role of

varieties in determining the observed patterns, the study relied on a qualitative analysis of untypical uses. The syntactic usage type was shown to contain a number of untypical-looking tokens that are affected by the broader lack of grammaticality in the utterance. The highest percentage of untypical uses came from the situational usage type. These untypical tokens could be arranged on a cline according to their level of identifiability. Those nearest to the unacceptable end were the least identifiable. Speakers who use these referents only indicate possession of their knowledge without identifying them to the listener. No untypical uses were found in the other usage types.

Findings similar to *the* were obtained for *a/an* in the quantitative analysis. Variation was again shown to be more amenable to text types than varieties. The qualitative analysis indicated that there were a number of significant differences in the varieties. However, overall, no clear patterns existed for either the Inner Circle or the Outer Circle group. Some untypical uses were discovered in the Outer Circle varieties. The first-mention/quantitative usage type showed that a few tokens used *a/an* instead of *the*. These were thus construed as inclusive, implying that other similar referents were available. In the same usage type, it was shown that a number of tokens presupposed a recategorisation of number. This was due to the use of *a/an* with referents that would typically be treated as mass. There were no untypical uses found in the other usage types.

As regards \emptyset , the untypical tokens of undetermined 'house' NPs were mostly occurrences with the null article. This rendered the NPs identifiable, inclusive and

maximally abstract. For many referents, the null article turned them into abstract ideas or concepts. At times, these concepts were actually one-member sets which were otherwise proper names. Other than this, ‘house’ was found to be used in *in house* by way of misanalysis. This was due to ‘house’ in the adjectival compound ‘in + house’ being treated as a noun. Broadened in its scope of description, the phrase was consequently used as an adjective, an adverb and a PP. Finally, two untypical tokens were the result of broader ungrammaticality in the utterances.

The findings of this study add further depth to our understanding of articles and their patterns of use in the varieties. First of all, while they are shown to influence usage types, there is only a weak trend identified among the varieties i.e. Outer Circle varieties very slightly prefer the use of textual definites. In the bigger picture, however, this is not a significant finding because, as it does in a few other comparisons, AmE defies the (decreasing) trend of other Inner Circle varieties, which is also a trend followed by IndE. The trends of the varieties, collective or individual, are so unpredictable that in the end there is no evidence of an overall trend. The one trend that can be said to be constant in all the varieties is idiosyncrasy. This claim is consistent with the observations made in chapter five in which we analysed the raw frequencies of articles in the corpora. As with the conclusion drawn in that chapter, we must again submit that variation of article use in the varieties is a subtle phenomenon. However the analysis in this chapter has allowed us to describe in some detail its nature.

Firstly variation in quantitative terms in the Outer Circle is minimal. This actually runs counter to the claims made by some surveyors of linguistic features in world Englishes who claim that article variation is one of the most widespread morpho-syntactically variable items found worldwide (e.g. Kortmann and Szmreczanyi, 2004). Our findings show that the variation is most evident only in private conversation in which language samples of lower lects are also found. Conversely where a gatekeeping or editorial function is maintained, the variability is kept to a strict minimum, thereby mirroring the usage patterns of the Inner Circle. Collectively, spoken language of the informal kind only forms a part of a variety. When other text types are considered, the variability does diminish in its significance.

Having said this, there are distinct variation patterns that can be found in the Outer Circle varieties and we saw these in the previous three sections. When article variation occurs, however, the untypical quality of the utterances does leave an impression on the observer, which could possibly explain why it is frequently cited in such surveys.

Variation in the Inner Circle group itself is minimal and this is in no small part due to the data that have been used for the study. If more data are analysed, such variation will certainly be found.

In the previous chapter it was claimed that it is the interaction between the meaning of articles and the nouns they modify that determines the variation we are interested in. Considering the case study of 'house' this seems to be the case. We see now that

although 'house' is a common noun and has a particular dictionary meaning as a noun, it can occur with *a*, *the* or null. In other words, the fact that it is a common noun alone and/or its lexical semantic features alone do not determine the kind of article it takes. It is the speaker who chooses the article to construe the meaning that s/he wishes to express through the NP. For many of the untypical uses we identified, these meanings were examined with Chesterman's theory of definiteness.

In the final chapter, we will synthesise these and other findings in the thesis.

Chapter 8 Conclusion

The research began with three questions about article variation in Inner and Outer Circle varieties of English. This chapter concludes the thesis by synthesising the findings of the three studies conducted in the previous three chapters. It also aims to provide the answers to those research questions.

Section 8.1 presents a synthesis of the findings of the frequency analysis in chapter five, the collocational analysis in chapter six and the usage type study in chapter seven. 8.2 addresses the three research questions. 8.3 discusses the limitations of the study and presents recommendations for future research.

8.1 A synthesis

The overarching study described in this study was conducted in a top-down manner. The analysis began with the most obvious aspect of a corpus i.e. frequency. After looking at the absence or presence of articles in the text categories and varieties, we proceeded to examine the patterns that were manifested in their collocation, arguably a corpus property which is less obvious than the last one. Finally we delved further into their usage patterns by annotating the articles with information that was implicit on the surface.

In this synthesis we piece together the findings by going bottom-up i.e. beginning with what we understand about the usage types of articles through our observation of their collocational patterns to their actual manifestation in the corpora.

8.1.1 *The*

The usage type analysis shows that text types have a greater influence on the use of the definite article than do the varieties. Reportage, for example, is found to favour textual, structural and situational text types. Private conversation and academic writing each has a different preference. While it is certainly worthwhile to extrapolate from these findings, in terms of varieties, these are only tendencies which translate into idiosyncratic overall preferences for usage types. The most striking thing about these preferences is the absence of evidence that certain usage types are more prevalent in either group of varieties. This means that untypical usage patterns that are identified in Sand (2004) do not give rise to differential percentages of use in the Outer Circle varieties. This is also an observation made by Sand herself. How does the use of the article vary, then?

If the tokens of each usage type are examined further, we can detect a number of untypical, or marked, patterns of use in the Outer Circle varieties. Only one was found in the Inner Circle group. It is in these untypical usage patterns that we see the highest level of variation.

The situational text type produces the highest number of untypical usages, followed by the structural text type. The others did not produce any untypical tokens in the data. Untypical situational definites are mostly found in the spoken data and their variation can be attributed to the degree of their identifiability. Some referents of situational NPs are less identifiable than others. To illustrate this continuum, we place the NPs on a cline of acceptability. The less identifiable they are, the more unacceptable to the hearer they become. Only the Outer Circle varieties are shown to include the less identifiable referents in their NPs.

Untypical usages in the structural text type are also found in the spoken data only. Although smaller in number, their occurrence is caused by more than one factor. The first is ungrammaticality and the second is indiscriminate definiteness rule application. The latter is evidenced by *the disco [that is] there* which is evoked by the mention of a government ministry (see 183 in chapter seven).

The presence of both typical and untypical usages in the Outer Circle varieties and the general absence of the untypical ones in the Inner Circle varieties lead us to question their linguistic correlations, i.e. given these variation patterns, what do they show as corresponding patterns on the surface? The findings of the collocational analysis indicate to us that no particular kind of nouns appears as the most frequent collocate of the definite article. One significant exception to this claim is *the society*. In fact the use of general nouns, which is shown to differ between ESL and native speakers in several studies, is not indicative of any disparity between the two groups of varieties. The use

of specific nouns, on the other hand, is widespread in each variety with those in the Outer Circle demonstrating a greater variety. The high socio-cultural content in many of these nouns indicates a considerable degree of code-mixing in the Outer Circle. On the whole, however, variation in the usage types is undeniably imperceptible on the surface, underscoring the subtle semantic differences involved in it.

What the collocations of *the* can more concretely indicate to us is that the definite article is complemented in distinct ways in several groupings of varieties. This can be seen most clearly in its most frequent clusters. Many clusters with *the* in Inner Circle varieties are found to take different prepositions from those in the Outer Circle ones. Often the Inner Circle group is more flexible in its complementation patterns, creating more opportunities for the use of the article. The same level of homogeneity that is shared by the Inner Circle varieties has also led them to share a larger number of clusters between them. In contrast, the Outer Circle varieties not only contrast with their Inner Circle counterparts but additionally with each other, resulting in each a larger unique set of clusters than any of those produced by the Inner Circle varieties.

So far the varieties' collocations have been the most substantial evidence of their variation. Once again the subtlety of the variation is encountered in our analysis because the collocational patterns are not visible to the naked eye and these patterns do not represent radical changes in the grammar of articles.⁶²

⁶² But it could be argued that for some speakers, the different collocational patterns may suggest to them that there are subtle differences.

The findings reported in chapter seven show that text types are more influential than varieties in determining the usage types of the definite article. The findings presented in chapter six indicate that varieties have their own preferences in their complementation patterns of the definite article. These two sets of findings help to explain the observation we made in the frequency analysis in chapter five. Recall that private conversation makes use of the lowest percentage of *the* in all varieties except AmE. Spoken language, as we discovered in the last chapter, contains a number of untypical usage patterns, many of which involve the use of *a/an* or \emptyset instead of *the*. Although the actual number is small, we need to bear in mind that the analysis was conducted on only seven files of the text type. At the same time, we have to allow the possibility of speakers using other determinatives altogether or simply imprecise language, as was claimed in chapter five. Other text types are affected in other ways. While we have to admit that there must exist other factors not considered in the study, the varieties' collocational preferences for different complementation patterns are likely to have played a part. The same reason could partly explain why the varieties were each found to use different percentages of *the* in different text types as well.

8.1.2 *A/an*

In the case of *a/an*, the high number of untypical usages in private conversation involving *the* and \emptyset similarly has led to a lower occurrence of this definite article in private conversation. The findings of chapter six, too, seem to be in line with the frequencies of *a/an* in the other text types. Recall from chapter five that the frequencies of this article were not significantly different in all the text types except for private

conversation. This finding is compatible with the collocational patterns observed for *a/an*. In chapter six it was discovered that although the varieties generally had different preferences for the complementation patterns of *a/an*, their occurrences did not vary greatly in number. Often the difference between the frequencies of clusters from different varieties is one or two. Consequently, despite the different collocational preferences that affected the production of *the* were also present in the case of *a/an*, they did not have the number to substantially influence the production of the latter.

8.1.3 Ø

The analysis conducted in chapter seven shows that the untypical usages of Ø arises from the omission of *a/an* and *the*. In both cases the omission results in the null form, as opposed to the zero article. This leads to the application of extensivity which renders the (otherwise count and singular) entity unbounded, and thus, abstract.

8.1.4 All articles

The synthesis of the findings for all articles is diagrammed in figure 8.1. The diagram presents the most general findings for each article at each level of the study. It also shows the two major independent variables that were investigated i.e. variety and text type. Reading the diagram from top to bottom allows the reader to interpret the findings as an exploration, or in the manner this study was conducted. Reading it from bottom to top gives the reader a synthetic perspective of the findings and an understanding of the frequency patterns that are visible on the surface.

Figure 8.1: A diagrammatic representation of findings and their interpretation

		exploration				
		<i>the</i>	<i>a/an</i>	∅		
Frequency	V	- low use in private conversation - no distinct trends although differences are statistically significant i.e. idiosyncratic	- low use in private conversation - no distinct trends across varieties and differences are not statistically significant	not applicable	←	T
	A					E
Collocation	R	- distinct collocational preferences for complementation patterns - more homogeneity in Inner Circle varieties	- distinct collocational preferences for complementation patterns although low in number - more homogeneity in both circles	not applicable	←	X
	I					T
Usage type	E	- no distinct trends across varieties - untypical usages in Outer Circle especially in spoken language	- no distinct trends across varieties - untypical usages in Outer Circle especially in spoken language	- untypical usages of ∅ <i>house</i> are null not zero	←	T
	T					Y
		S	'the'	'a/an'		∅

synthesis

8.2 Research questions

At the beginning of the thesis the following questions were presented:

1. In naturally occurring data, what are the similarities and differences in the use of articles in Inner and Outer Circle varieties of English?
2. How can the differences be explained?
3. Where applicable, how do the findings of Outer Circle varieties compare to those of SLA?

Each of these is addressed in a section of its own.

8.2.1 Article use: similarities and differences between Inner and Outer Circles

Although not stated thus far, the similarities between the two groups and those shared by all individual varieties far outweigh the differences observed. If we measure the level of similarities against any accepted standard of grammar, the use of articles cannot be said to be widely divergent. In the data there is a large common core that is shared by both groups of varieties. Almost all of the usage types listed in table 2.1, for example, are found in the data. It is not impossible that in a larger set of data, all of them could be found in all varieties.

In the light of the statement that was made about their similarities, the differences that have been uncovered should not be described as ‘many’ but ‘varied’. There are several kinds and we will try to categorise them according to frequency, collocation or usage type.

8.2.1.1 Frequency differences

There are many differences in the frequencies of *the* and *a/an* in the varieties (see chapter five); we focus here on the major ones.

1. The Outer Circle, as represented by SgE, IndE and PhilE, is found have a higher tendency to use more or fewer definite articles in several text types. However the significance of this trend is seriously undermined by the inclusion of NZE among them, and to a lesser extent, the absence of KenE.
2. While all varieties (except AmE whose findings are interpreted with caution) are shown to use a less-than-average number of definite articles in private conversation, the Inner and Outer Circles are generally divided on its use in correspondence. The Outer Circle group's use parallels that of private conversation, i.e. low. However this is also found to be the trend of BrE.

In general, there are tendencies to use more or fewer definite articles in the Outer Circle group but their significance is weakened by the presence of one Inner Circle variety, usually NZE or AmE, or the absence of one of their own, usually KenE.

8.2.1.2 Collocational differences

The major collocational differences between the two groups of varieties are:

1. There are more unique clusters in the Inner Circle group and many of these are shared by at least three of the varieties. The Outer Circle varieties do not share any cluster among them.
2. Clusters in the Inner Circle group tend to employ more types of complements

than do those in the Outer Circle group. Whereas these complements only vary in number across the Inner Circle varieties, they are often altogether absent in the Outer Circle ones.

3. In the Outer Circle varieties, there is a substantial amount of code-mixing in the collocations of *the* but not *a/an*. Socio-cultural content appears to be an important factor in definiteness marking.
4. *The society* is a significant collocation in the Outer Circle group.

8.2.1.3 Usage type differences

The most significant difference between the usage types in the two groups of varieties is the presence of untypical patterns of use in the Outer Circle varieties. The aberration is sometimes due to ungrammaticality but more often is due to meaning.

1. In the Outer Circle group's situational definites, some referents are found to be less identifiable than others. This is hardly the case for the Inner Circle varieties.
2. There is limited evidence of a tendency to mark NPs with *the* (instead of *a*) if it is postmodified. This is referred to as 'indiscriminate definiteness rule application'.

8.2.2 Explaining the differences

The differences that are listed above have been analysed in detail in their respective chapters (see also the synthesis presented above). In this section we expound on the role of the semantics of definiteness in giving rise to the frequencies, patterns and usages of

articles that we have observed in this study. While we have certainly enriched our understanding of definiteness by drawing on several different theories (see chapter two), the most suitable theory to explain the variation being observed is Chesterman (1991/2005). To enable all three articles (and the null form) to be discussed, we shall use ‘house’ for illustration where necessary. As observed by Chesterman definiteness is a composite of three meanings: locatability, inclusiveness and extensivity (see 2.5.3 and table 2.3). Each article has a different composition of meanings.

When a speaker says ‘I have a house’ we interpret it as a first-mention/quantitative. When another says ‘I see the house’ we probably have already been engaged in the discourse and perhaps interpret the referent as textual or situational. When yet another says ‘I buy Ø houses’, we tend to think of the plural referent in the same way we do for ‘a house’. These are the typical interpretations that we give the utterances. When a speaker utters *Twenty-two kilometres from Ø house* (see 244), however, we have to evoke the notion of extensivity so that this entity can be perceived, albeit as an idea or a concept (see 7.5). Based on the findings, it is possible that extensivity plays a significant role in the use of articles with singular nouns in the Outer Circle varieties.

What we have not yet considered is the fact that any of these modified patterns of use can be the result of a reconfiguration of the three composite meanings. Let us examine the following substitutions. The substituted article is placed before the arrow while the substituting article occurs after it:

1. *a/an* → *the*

e.g. *the cool climate* in (186)

Meaning: {not locatable, not inclusive, not extensive} → {locatable, inclusive, not extensive}

Interpretation: one of known types of climates, possibly one out of a set of two {warm, cool}.

2. *a/an* → ∅ (null)

e.g. ∅ *terrace house* (235)

Meaning: {not locatable, not inclusive, not extensive} → {locatable, inclusive, extensive}.

Interpretation: an unbounded entity, an abstract idea or concept of '(terrace) house'.

3. ∅ (null) → *the*

e.g. *the society* (196) – (198)

Meaning: {locatable, inclusive, extensive} → {locatable, inclusive, not extensive}

Interpretation: the society that is of interest (cf. She is *the Sophia Loren*).

4. ∅ (zero) → *a/an*

e.g. *a tomato sauce* (213)

Meaning: {not locatable, not inclusive, extensive} → {not locatable, not inclusive, not extensive}

Interpretation: a kind of tomato sauce, but more likely, an unspecified amount of tomato sauce.

5. *the* → *a/an*

e.g. *an extent of tearing* (204)

Meaning: {locatable, inclusive, not extensive} → {not locatable, not inclusive, not extensive}

Interpretation: there exist other extents of tearing.

6. *the* → ∅ (null)

e.g. *twenty-two kilometres from ∅ house* (224)

Meaning: {locatable, inclusive, not extensive} → {locatable, inclusive, extensive}

Interpretation: an unbounded entity, an abstract idea or concept of 'house'.

It is important to note that reconfiguration can also occur in the Inner Circle varieties. One oft-quoted example is 'hospital' in AmE's 'take me to the hospital' and BrE's 'take me to hospital'. Using the same analysis used above, we can say that the British use is a composite of {locatable, inclusive, extensive} as a result of the null article while the American use is a composite of {locatable, inclusive, not extensive} due to the definite article. In our data, the most conspicuous case of reconfiguration in Inner Circle varieties is the use of *∅ house* in NZE in (254).

Reconfigured meanings such as the above are most clearly seen in the study of usage types, where the examples are taken from. However it is not difficult to imagine the effects they have on the collocational patterns, although our focus on the most frequent patterns will not uncover them. Finally, the reconfiguration of these meanings definitely affects the choice of one article over another. On the surface these choices are naturally

reflected in their frequencies. This is another reason why it is hard to observe any trends in the frequencies in the varieties.

8.2.3 Nativisation vs. SLA

The areas in which comparison can be made with SLA are (1) L1 influence (2) proficiency and (3) specificity. These are discussed in turn in the following sections. Finally the discussion focuses on nativisation.

8.2.3.1 L1 influence

Overall, it is difficult to say that the absence or presence of articles in the L1 has affected the frequencies of articles. This is shown by the frequencies obtained in Phile (see chapter five for a detailed discussion). Effects on the collocational patterns have also been difficult to detect. On the other hand, very few untypical usages coming from this variety were found in the usage type analysis. This does suggest some degree of positive transfer from its L1.

Based on the evidence presented, it would be hasty to conclude that the presence or absence of articles in the L1 has no effect on article use. However, we have to concede that whatever effect it may have is not fully manifest in our analysis. Conversely we may claim that such effects are minimal in naturally occurring data.

With other aspects of L1 transfer, we have to be similarly cautious in stating the effects that L1 may have on article use. For example, although the different complementation patterns in the Outer Circle varieties may be the result of L1-related effects, we also have to acknowledge that other factors such as simplicity may play a role (see external vs. internal based factors in 3.1).

8.2.3.2 Proficiency

In chapter five we saw that there is a suggestion of proficiency playing a role on the production of the definite article based on the low frequencies in private and conversations (see 5.5). Despite this, we also acknowledge the possible influence of level of formality, which later proved to be more definitive in explaining the results. However it needs to be borne in mind that less proficient samples of English do not constitute the majority of the data at all. More substantial evidence of the influence of proficiency in the use of articles came from the usage type analysis. In chapter seven we saw that a few untypical tokens of *the*, *a/an* and \emptyset that were affected by the lack of grammaticality in the speakers' utterances. However, these were few and far between. While the collocational analysis did not offer any corroboration, the analysis showed that Outer Circle varieties' speakers do not overuse general nouns as collocates, a practice otherwise associated with ESL learners. Therefore we have sufficient evidence to say that in naturally occurring data of Outer Circle varieties, proficiency of speakers only minimally affects the use of articles.

This claim is in accordance with the distinction that we seek to maintain between nativisation and interlanguage in Outer Circle varieties. The unstable nature of interlanguage grammar was found to affect fewer than five NPs. The other untypical usages were not affected by ungrammaticality but by change in their meanings. (The ungrammaticality is usually evident in other parts of the utterance too.) In the case of \emptyset *house*, it is possible to work out the resulting change in its meaning. This points to the existence of nativisation.

8.2.3.3 Specificity

In conformity with many SLA studies (see 7.3.3.1 and 2.6), this study discovered that [\pm specificity] is a feature that considerably influences article use in the Outer Circle varieties. Figure 7.4 (chapter seven) indicates the extent of its influence on definite NPs of this group. All untypical usages of situational definites are based on specificity. The closer a referent is placed towards the ‘unacceptable’ end of the cline, the more specific it becomes.

8.2.3.4 Nativisation

Three notable features of nativisation are identified in the study. These are (1) the use of specificity as the basis of definiteness; (2) the application of extensivity; and (3) recategorisation of countness. The first is found in the insertion of the definite article. The second occurs in the omission of *the* and *a/an*. The third is the result of *a/an* insertion.

In the light of our discussion of nativisation in chapter three, there are a few additional observations that can be made regarding the above features. The use of specificity in *the*-insertion and the recategorisation of countness in *a/an*-insertion can be said to be instances of reanalysis. This is because speakers create new patterns of use (or forms) to accommodate the meanings that they want to express. The application of extensivity, on the other hand, is an instance of simplification. This is because of the perceived simplification that takes place when the article is omitted, which in turn suggests that the marking of (non)identifiability is unnecessary. Consistent with Schneider's claim that the new practice may actually run counter to the perceived simplicity of the new form (see 3.1.1), speakers who apply extensivity in effect require their listeners to rely heavily on pragmatics to disambiguate the intended referent.

Now that the innovations are identified, the next task is to describe their diffusion. It appears that this process is far from straightforward in the Outer Circle varieties. All three patterns of use are available in the feature pool of the varieties. However, given what we know about the use of these patterns in more formal text types as opposed to the highly informal private conversation, further diffusion of these innovations seems to be restrained by language gate-keeping agents. Meanwhile the continuing presence of these usages in informal language indicates that they have been selected for this register, both for social (e.g. peer acceptance) and functional (e.g. reduction of rule complexity) reasons (see socially based vs. functionally based selection in 3.1.2).

8.3 Limitations of the study and recommendations for future research

While this study has been premised on the naturalness of corpus data, the results indicate that the perspective gained here can be considerably enriched by data provided by elicitation. It is clear to see that although the corpus-based investigation has been successful in describing all the patterns of use observed in the analysed data, many more uses that are described in reference grammars are not attested.

This brings us to the second limitation of corpus data. The unavailability of a comparable ICE component for AmE has really complicated the analyses. For a corpus-based comparison of world Englishes to be comprehensive, the complete American data are urgently needed.

Future research obviously needs to include both types of data in order to ensure that all uses of articles are covered. By doing so, all known uses can be attested and most, if not all, possible uses can be encountered.

Finally to further our understanding of article use in the varieties, future research might include in its analysis text types that were not analysed in this thesis.

Bibliography

- Abbott, B. (2006) 'Definite and indefinite', in K. Brown (ed) *The Encyclopedia of Language and Linguistics, 2nd ed, vol. 3*, Oxford: Elsevier, pp. 392-399.
- Abbott, B. (2004) 'Definiteness and indefiniteness', in L. Horn and G. Ward (eds) *Handbook of Pragmatics*, Oxford: Blackwell, pp. 122-149.
- Abbott, B. (2001) 'Definiteness and identification in English', in N. Enikö (ed) *Pragmatics in 2000: Selected papers from the 7th International Pragmatics Conference, vol. 2*, Antwerp: International Pragmatics Association, pp. 1-15.
- Abbott, B. (1999) 'The formal approach to meaning', *Journal of Foreign Languages* 119: 2-20.
- Abdulaziz, M. (1991) 'East Africa (Tanzania and Kenya)', in J. Cheshire (ed) *English Around the World: Sociolinguistic Perspectives*, Cambridge: Cambridge University Press, pp. 391 – 491.
- Algeo, J. (2006) *British or American English? A Handbook of Word and Grammar Patterns*, Cambridge and New York: Cambridge University Press.
- Ansaldo, U. (2004) 'The evolution of Singapore English: finding the matrix', in L. Lim (ed) *Singapore English: A Grammatical Description*, Amsterdam and Philadelphia: John Benjamins, pp. 127 - 149.
- Ariel, M. (1990) *Accessing Noun-phrase Antecedents*, London: Routledge/Croom Helm.
- Ariel, M. (1988) 'Referring and accessibility', *Journal of Linguistics* 24: 65–87.

- Baayen, R. (2008) *Analyzing Linguistic Data*, Cambridge and New York: Cambridge University Press.
- Bach, K. (1999) 'The myth of conventional implicature', *Linguistics and Philosophy*. 22: 327-366.
- Baker, P., Hardie, A. and McEnery, T. (2006) *A Glossary of Corpus Linguistics*, Edinburgh: Edinburgh University Press.
- Bamgbose, A. (1998) 'Torns between the norms: innovations in world Englishes', *World Englishes* 17: 1-14.
- Bamgbose, A. (1995). 'English in the Nigerian Environment', in Bamgbose et al. (eds) *New Englishes: A West African Perspective*, Ibadan: Mosuro. pp. 9-26.
- Behrens, L. (2005) 'Genericity from a cross linguistic perspective', *Linguistics* 43: 275-344.
- Berry, R. (1991) 'Re-articulating the articles', *English Language Teaching Journal* 45: 252-9.
- Bhatt, R. (2001) 'World Englishes', *Annual Review of Anthropology* 30: 525-550.
- Biber, D. and Conrad, S. (2001) 'Register variation: A corpus approach', in D. Schiffrin, D. Tannen, and H. Hamilton (eds), *The Handbook of Discourse Analysis*, Oxford: Blackwell, pp. 175-96.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman Grammar of Spoken and Written English*, Harlow: Longman.
- Bickerton, D. (1981) *Roots of language*, Ann Arbor: Karoma.
- Birner, B. and Ward, G. (1998) *Information Status and Noncanonical Word Order in English*, Amsterdam: John Benjamins.

- Birner, B. and Ward, G. (1994) 'Uniqueness, familiarity, and the definite article in English', in S. Gahl et al. (eds) *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society*, Berkeley: Berkeley Linguistics Society, pp. 93-102.
- Bruthiaux, P. (2003) 'Squaring the circles: Issues in modeling English worldwide', *International Journal of Applied Linguistics*, 13: 159-178.
- Burridge, K. (2004) 'Synopsis: morphological and syntactic variation in the Pacific and Australasia', in B. Kortmann et al. (eds) *A Handbook of Variation of English, Vol. 2: Morphology and Syntax*, Berlin and New York: Mouton de Gruyter, pp. 1116-1131.
- Burridge, K. and Kortmann, B. (2004) 'Introduction: varieties of English in the Pacific and Australasia', in B. Kortmann et al. (eds.) *A Handbook of Variation of English, Vol. 2: Morphology and Syntax*, Berlin and New York: Mouton de Gruyter, pp. 547-559.
- Burton-Roberts, N. (1981) 'Review of J. Hawkins, Definiteness and indefiniteness', *Language* 57: 191-6.
- Burton-Roberts, N. (1976) 'On the generic indefinite article', *Language* 52: 427-48.
- Bybee, J. (1994) 'The grammaticization of zero: Asymmetries in tense and aspect systems', in W. Pagliuca (ed) *Perspectives on Grammaticalization*, Amsterdam: John Benjamins, pp. 235-54.
- Carlson, G. and Sussman, R. (2005) 'Seemingly indefinite definites', in S. Kessler and M. Reis (eds) *Linguistic Evidence: Empirical, Theoretical, and Computational perspectives*, Berlin and New York: Mouton de Gruyter, pp. 71-85.

- Carlson, G. and Pelletier, F. (eds) (1995) *The Generic Book*, Chicago: University of Chicago Press.
- Carston, R. (2002). *Thoughts and Utterances. The Pragmatics of Explicit Communication*, Oxford: Blackwell.
- Carter, R. and McCarthy, M. (1997) *Exploring Spoken English*, Cambridge: Cambridge University Press.
- Chesterman, A. (1991/2005) *On Definiteness: a Study with Special Reference to English and Finnish*, Cambridge: Cambridge University Press.
- Christophersen, P. (1939) *The Articles: A Study of their Theory and Use in English*, Copenhagen: Munksgaard.
- Clark, H. (1977) 'Bridging', in P. Johnson-Laird and P. Watson (eds) *Thinking: Readings in Cognitive Science*, Cambridge: Cambridge University Press, pp. 169-174.
- Clark, H. and Haviland, S. (1977) 'Comprehension and the given-new contract', in Freedle, R. (ed) *Discourse Production and Comprehension*, Norwood: NJ: Ablex, pp.1-40.
- Collins, P. (2004) 'Let-imperatives in English', *International Journal of Corpus Linguistics* 9: 299-319.
- Collins, P. (1998) *English Grammar*, Melbourne: Longman.
- Collins, P. (1991) *Cleft and Pseudo-Cleft Constructions in English*, London and New York: Routledge.
- Collins, P. and Peters, P. (2004) 'Australian English: Morphology and syntax', in B. Kortmann et al. (eds.) *A Handbook of Variation of English*, Vol. 2:

- Morphology and Syntax*, Berlin and New York: Mouton de Gruyter, pp. 601–10.
- Croft, W. (2000) *Explaining Language Change*, London and New York: Longman.
- Crystal, D. (2006) 'English worldwide', in R. Hogg and D. Denison (eds) *A History of the English Language*, Cambridge: Cambridge University Press, pp. 420-39.
- Davies, M. (2004) *BYU-BNC: The British National Corpus*, Available online at <http://corpus.byu.edu/bnc>.
- Davies, M. (2009) *The Corpus of Contemporary American English (COCA)*, Available online at <http://www.americancorpus.org>.
- Declerck, R. (1986) 'Two notes on the theory of definiteness', *Journal of Linguistics* 22: 25-39.
- De Klerk, V. (2005) 'Procedural meanings of "well" in a corpus of Xhosa English', *Journal of Pragmatics* 37: 1183-1205.
- Dixon, R. (2006) 'The articles in English', *Studia Anglica Posnaniensia* 42: 31-46.
- Donnellan, K. (1966) 'Reference and definite descriptions', *Philosophical Review* 75: 281-304.
- Downing, A. and Locke, P. (2000) *A University Course in English Grammar*, London and New York: Routledge.
- Du Bois, J. (1980) 'Beyond definiteness: The trace of identity in discourse', in W. Chafe (ed) *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, Norwood: NJ: Ablex, pp. 203-274.
- Epstein, R. (2002) 'The definite article, accessibility, and the construction of discourse referents', *Cognitive Linguistics* 12: 333-378.

- Epstein, R. (1999) 'Roles, frames and definiteness', in K. Van Hoek, A. Kibrik and L. Noordman (eds) *Discourse Studies in Cognitive Linguistics*, Amsterdam: John Benjamins, pp. 53-74.
- Epstein, R. (1992) *Discourse and Definiteness: Synchronic and Diachronic Perspectives*, PhD Dissertation, San Diego: University of California.
- Farrar, K. and Jones, M. (2002) 'Introduction', in M. Jones and E. Esch (eds) *Language Change: The Interplay of Internal, External and Extra-linguistic Factors*, Berlin and New York: Mouton de Gruyter, pp.1-16.
- Fauconnier, G. (1998) 'Mental spaces, language modalities, and conceptual integration', in M. Tomasello (ed) *The New Psychology of Language*, Mahwah, NJ: Lawrence Erlbaum, pp. 251-279.
- Fauconnier, G. (1985) *Mental Spaces: Aspects of Meaning Construction in Natural Language*, Cambridge, MA: Massachusetts Institute of Technology Press.
- Fennell, B. (2007) 'Colonial and postcolonial varieties', in C. Llamas, L. Mullany and P. Stockwell (eds) *The Routledge Companion to Sociolinguistics*, New York: Routledge, pp. 192-198.
- Filppula, M. (1999) *The Grammar of Irish English*, London. Routledge.
- Fodor, J. and Sag, I. (1982) 'Referential and quantificational indefinites', *Linguistics and Philosophy* 5: 355-398.
- Foley, J. (1998) 'Code-switching and learning among young children in Singapore', *International Journal of the Sociology of Language*, 130: 129-150.
- Fong, V. (2004) 'The verbal cluster', in L. Lim (ed) *Singapore English: A Grammatical Description*, Amsterdam and Philadelphia: John Benjamins, pp.

75-99.

Fraurud, K. (1990) 'Definiteness and the processing of NPs in natural discourse',
Journal of semantics 7: 395-433.

Garcia Mayo, M. (2008) 'The acquisition of four nongeneric uses of the article *the* by Spanish EFL learners', *System*, 36: 550-565.

Gil, D. (1994) 'Genitives, number and (in)definiteness: Some data from English, Singlish, Mandarin and Malay', in F. Planck (ed) *Agreement Gender Number Genitive (EUROTYP Working Papers VII/23)*, Berlin: The European Science Foundation, EUROTYP Programme, pp. 109 -131.

Givón, T. (1984) *Syntax: A Functional-Typological Introduction, vol 1*, Amsterdam: John Benjamins.

Givón, T. (1992) 'On interpreting text-distributional correlations: some methodological issues' in D. Payne (ed) *Pragmatics of Word Order Flexibility*, Amsterdam: John Benjamins, pp. 305-315.

Goto Butler, Y. (2002) 'Second language learners' theories on the use of English articles', *Studies in Second Language Acquisition* 24: 451-480.

Granger, S. (1998) *Learner English on computer*, London: Longman.

Grant, L. and Bauer, L. (2004) 'Criteria for re-defining idioms: Are we barking up the wrong tree?', *Applied Linguistics* 25: 38-61.

Greenbaum, S. (1996) 'Introducing ICE', in S. Greenbaum (ed) *Comparing English Worldwide: The International Corpus of English*, Oxford: Clarendon Press, pp. 3-12.

- Gundel, J., Hedberg, N. and Zacharski, R. (1993) 'Cognitive status and the form of referring expressions in discourse', *Language*. 69: 274-307.
- Haegeman, L. and Gueron, J. (1999) *English Grammar. A Generative Perspective*, Oxford: Blackwell
- Halliday, M. A. K. and Hasan, R.(1976) *Cohesion in English*, London: Longman.
- Hawkins, R. (2001) *Second Language Syntax: A Generative Introduction*, Oxford and New York: Blackwell.
- Hawkins, R and Chan, C. (1997) 'The partial availability of Universal Grammar in second language acquisition: the failed functional features hypothesis', *Second Language Research* 13: 187-226.
- Hawkins, J. (1978) *Definiteness and Indefiniteness*, London: Croom Helm.
- Hawkins, J. (1991) 'On (in)definite articles: implicatures and (un)grammaticality prediction', *Journal of Linguistics* 27: 405-442.
- Heine, B. (1997) *Cognitive Foundations of Grammar*, Oxford and New York: Oxford University Press.
- Hickey, R. (2007) *Irish English: History and Present-day Forms*, Cambridge and New York: Cambridge University Press.
- Huddleston, R. and Pullum, G. (2005) *A Student's Introduction to English Grammar*, Cambridge: Cambridge University Press.
- Huddleston, R. and Pullum, G. (2002) *The Cambridge Grammar of the English Language*, Cambridge: Cambridge University Press.
- Hudson-Eattle, D. and Schmied, J. (1999) *Manual to Accompany the East African Component of The International Corpus of English*, Chemnitz: Chemnitz

University of Technology.

Huebner, T. (1983) *A Longitudinal Analysis of the Acquisition of English*, Ann Arbor: Caroma.

Hunston, S. and Francis, G. (1999) *Pattern Grammar*, Amsterdam: John Benjamins.

Ionin, T., Ko, H. and Wexler, K. (2004) 'Article semantics in L2-acquisition: the role of specificity', *Language Acquisition* 12: 3-69.

Ionin, T., and Wexler, K. (2003) 'The certain uses of the in L2-English', in J. Liceras, H. Zobl and H. Goodluck (eds) *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference*, Somerville, MA: Cascadilla Press, pp.150-160.

Irving, K. (1984) 'Cross-cultural awareness in the ESL classroom', *Theory into Practice* 23: 138-143.

Jarvis, S. (2002) 'Topic continuity in L2 article use', *Studies in Second Language Acquisition* 24: 387-418.

Kachru, B. (1992) 'Teaching world Englishes', in B. Kachru (ed) *The Other Tongue*, Urbana: University Illinois Press, pp. 91-107.

Kachru, B. (1986) *The Alchemy of English: The Spread, Functions and Models of Non-native Englishes*, Oxford: Pergamon.

Kachru, B. (1985) 'Standards, codification and sociolinguistic realism: The English language in the outer circle', in R. Quirk and H. Widdowson (eds) *English in the World: Teaching and Learning the Language and Literatures*, Cambridge: Cambridge University Press, pp. 11-36.

- Kachru, B. (1983) *The Indianization of English. The English Language in India*, Oxford: Oxford University Press.
- Kachru, B. (1982) 'Models for non-native Englishes', in B. Kachru (ed) *The Other Tongue*, Urbana: University of Illinois Press, pp. 48-70.
- Kerswill, P. and Williams, A. (2002) "'Salience" as an explanatory factor in language change: evidence from dialect levelling in urban England', in M. Jones and E. Esch (eds) *Language Change. The Interplay of Internal, External and Extra-linguistic Factors*, Berlin and New York: Mouton de Gruyter, pp. 81-110.
- Kioko, A. and Muthwii, M. (2004) 'English variety for the public domain in Kenya: speaker's attitudes and views', in A. Kioko and M. Muthwii (eds) *New Language Bearings in Africa: A Fresh Quest*, Clevedon: Multilingual Matters, pp. 34-49.
- Klinge, A. (2006) 'The unitary procedural semantics of "the", "this" and "that"', in *Acta Linguistica Hafniensi* 38: 54-77.
- Kortmann, B. (2006) 'Syntactic variation in English: A global perspective', in B. Aarts and A. McMahon (eds) *Handbook of English Linguistics*, Malden, MA, Oxford and Carlton, Vic.: Blackwell, pp. 603-624.
- Kortmann, B. and Szmreczanyi, B. (2004) 'Global synopsis: morphological and syntactic variation in English', in B. Kortmann, et al. (eds) *A Handbook of Varieties of English, Vol. 2: Morphology and Syntax*, Berlin and New York: Mouton de Gruyter, pp. 1142-1202.
- Kortmann, B., Burrige, K., Mesthrie, R., Schneider, E. and Upton, C. (eds) (2004) *A*

- Handbook of Variation of English, Vol. 2: Morphology and Syntax*, Berlin and New York: Mouton de Gruyter.
- Kripke, S. (1977) 'Speaker's reference and semantic reference', in P. French et al. (eds) *Midwest Studies in Philosophy Volume II: Studies in the Philosophy of Language*, Morris, MN: University of Minnesota Press, pp. 255-276.
- Lardiere, D. (2003) 'The comparative fallacy revisited: A reply to Lakshmanan and Selinker', *Second Language Research* 19: 129-143.
- Lakoff, G. (1982) *Categories and Cognitive Models*, Trier: Linguistic Agency University Trier.
- Lee, J. and Collins, P. (2004) 'On the usage of "have", "dare", "need", "ought" and "used to" in Australian English and Hong Kong English', *World Englishes* 23: 501-513.
- Leech, G. (1983) *Principles of Pragmatics*, London: Longman.
- Lewis, D. (1979) 'Scorekeeping in a language game', *Journal of Philosophical Logic* 8: 339-359.
- Lim, L. (2007) 'Mergers and acquisitions: On the ages and origins of Singapore English particles', *World Englishes* 26: 446-473.
- Lim, L. and Foley, J. (2004) 'English in Singapore and Singapore English', in L. Lim (ed) *Singapore English: A Grammatical Description*, Amsterdam and Philadelphia: John Benjamins, 1-18.
- Liu, D. and Gleason, J. (2002) 'Acquisition of the article *the* by nonnative speakers of English: An analysis of four nongeneric uses', *Studies in Second Language Acquisition*, 24: 1- 26.

- Löbner, S. (1985) 'Definites', *Journal of Semantics*, 4: 279-326.
- Lowenberg, P. (1990) 'Standards and norms for world Englishes: issues and attitudes', *Studies in the Linguistic Sciences* 30: 123-37.
- Lowenberg, P.(1986) 'Non-native varieties of English: nativization, norms, and implications', *Studies in Second Language Acquisition* 8:1-18.
- Ludlow, P. and Neale, S. (1991) 'Indefinite descriptions: In defence of Russell', *Linguistics and Philosophy* 14: 171-202.
- Lyons C. (1999) *Definiteness*, Cambridge: Cambridge University Press.
- Lyons, J. (1977) *Semantics*, Cambridge: Cambridge University Press.
- Mahlberg, M. (2005) *English General Nouns: A Corpus Theoretical Approach*, Amsterdam and Philadelphia: John Benjamins.
- Mahlberg, M. (2003). 'The textlinguistic dimension of corpus linguistics: The support function of English general nouns and its theoretical implications', *International Journal of Corpus Linguistics* 8: 97-108.
- Mair, C. (2002) 'Creolisms in an emerging: Written English in Jamaica', *English World-Wide* 23: 31-58.
- Master, P. (1997) 'The English article system: Acquisition, function and pedagogy', *System* 25: 215-232.
- Master, P. (1990) 'Teaching the English articles as a binary system', *TESOL Quarterly* 24: 461 – 478.
- Master, P. (1987) *A cross-linguistic Interlanguage Analysis of the Acquisition of the English Article System*, PhD dissertation, Los Angeles: University of California.

- McEnery, T and Wilson, A. (1996) *Corpus Linguistics*, Edinburgh: Edinburgh University Press.
- McMahon, A. (1994) *Understanding Language Change*, Cambridge University Press.
- Mesthrie, R. (2004) 'Introduction: varieties of English in Africa and Southeast Asia' in B. Kortmann et al. (eds.) *A Handbook of Variation of English, Vol. 2: Morphology and Syntax*, Berlin and New York: Mouton de Gruyter, pp. 805-812.
- Meyer, C. (2002). *English Corpus Linguistics: An Introduction*, Cambridge University Press.
- Milroy, J and Milroy, L. (1999) *Authority in Language: Investigating Standard English*, London and New York: Routledge.
- Mitchell, A. and Delbridge, A. (1965) *The Pronunciation of English in Australia*, Sydney: Angus and Robertson.
- Mollin, S. (2007) 'New variety or learner English? Criteria for variety status and the case of Euro-English', *English World-Wide* 28: 167-185.
- Mufwene, S. (2001) *The Ecology of Language Evolution*, Cambridge: Cambridge University Press.
- Neale, S. (1990) *Descriptions*, Cambridge, MA: Massachusetts Institute of Technology Press.
- Nelson, G. (1996) 'The Design of the Corpus' in S. Greenbaum (ed) *Comparing English Worldwide: The International Corpus of English*, Oxford: Clarendon Press, pp. 27-35.

- Nkemleke, D. (2006) 'Nativization of dissertation acknowledgements and private letters in Cameroon', *Nordic Journal of African Studies* 15: 166-184.
- Ooi, V. (2001) 'Globalising Singaporean-Malaysian English in an inclusive learner's dictionary', in B. Moore (ed) *Who's Centric Now? The Present State of Post-Colonial Englishes*, Oxford: Oxford University Press, pp. 95-121.
- Pakir, A. (1991) 'The range and depth of English-knowing bilinguals in Singapore', *World Englishes* 10: 167-179.
- Parrish, B. (1987). 'A new look at methodologies in the study of article acquisition for learners of ESL', *Language Learning* 37: 361-383.
- Partington, A. (1998) *Patterns and Meanings: Using Corpora for English Language Research and Teaching*, Amsterdam and Philadelphia: John Benjamins.
- Payne, T. (1997/2006) *Describing Morphosyntax*, Cambridge: Cambridge University Press.
- Platt, J. and Weber, H. (1980) *English in Singapore and Malaysia: Status, Features, Functions*, Kuala Lumpur: Oxford University Press.
- Poesio, M. and Vieira, R. (1998) 'A corpus-based investigation of definite description use', *Computational Linguistics* 24: 183-216.
- Prince, E. (1992) 'The ZPG letter: subjects, definiteness, and information status', in S. Thompson and W. Mann (eds) *Discourse Descriptions: Diverse analyses of a Fund-Raising Text*, Amsterdam: John Benjamins, pp. 295-325.
- Prince, E. (1981) 'On the inferencing of indefinite *this* NPs' in A. Joshi et al. (eds) *Elements of Discourse Understanding*, Cambridge: Cambridge University Press, pp.231-250.

- Prince, E. (1981) 'Toward a taxonomy of given-new information' in P. Cole (ed) *Radical Pragmatics*, New York: Academic Press, pp. 223-256.
- Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*, Harlow: Longman.
- Quirk, R. (1985) *English in the World*, Cambridge: Cambridge University Press.
- Ringbom, H. (1998) 'High-frequency verbs in the ICLE corpus' in A. Renouf (ed.) *Explorations in Corpus Linguistics*, Amsterdam: Rodopi, pp. 191-200.
- Robertson, D. (2000) 'Variability in the use of the English article system by Chinese learners of English', *Second Language Research* 16: 35-172.
- Romaine, S. (1998) 'Introduction', in S. Romaine (ed) *The Cambridge History of the English Language*, Cambridge and New York: Cambridge University Press, pp.1-55.
- Rosch, E., Mervis, C., Gray, W., Johnson, D. and Boyes-Braem, P. (1976) 'Basic objects in natural categories', *Cognitive Psychology* 8: 382-439.
- Rosenbach, A. (2008) 'Language change as cultural evolution: evolutionary approaches to language change', in Regine Eckardt, Gerhard Jäger and Tonjes Veenstra (eds) *Variation, Selection, Development: Probing the Evolutionary Model of Language Change*, Berlin and New York: Mouton de Gruyter, pp.23-72.
- Russell, B. (1905) 'On denoting', *Mind*, 14: 479-493.
- Sand, A. (2004) 'Shared morpho-syntactic features in contact varieties of English: article use', *World Englishes* 23: 281-298.

- Sankoff, G. and Mazzie, C. (1991) 'Determining noun phrases in Tok Pisin', *Journal of Pidgin and Creole Languages*, 6: 1-24.
- Schmied, J. (2004) 'East African English (Kenya, Uganda, Tanzania): morphology and syntax', in B. Kortmann et al. (eds) *A Handbook of Variation of English, Vol. 2: Morphology and Syntax*, Berlin and New York: Mouton de Gruyter, pp. 929-947.
- Schneider, E. (2007) *Postcolonial English*, Cambridge and New York: Cambridge University Press.
- Scott, M. (2004) *WordSmith Tools version 4*, Oxford: Oxford University Press
- Sharma, D. (2005) 'Language transfer and discourse universals in Indian English article use', *Studies in Second Language Acquisition* 27: 535-566.
- Sidner, C. (1978) 'The use of focus as a tool for disambiguation of definite noun phrases', in D. Waltz (ed) *Theoretical Issues in Natural Language Processing 2*, Urbana-Champaign, IL: University of Illinois at Urbana-Champaign, pp. 86-95.
- Sinclair, J. (1991) *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.
- Smith, A. (2009) 'Non-numerical quantifiers', in P. Peters, P. Collins and A. Smith (eds) *Comparative Studies in Australian and New Zealand English Grammar and Beyond*, Amsterdam and Philadelphia: John Benjamins, pp. 159-180.
- Startvik, J. and Leech, G. (2006) *English: One Tongue, Many Voices*, Basingstoke and New York: Palgrave Macmillan.
- Strawson, P. (1950) 'On referring', *Mind* 59: 320-344.
- Stubbs, M. (2007) 'Quantitative Data on Multi-Word Sequences in English: the Case

- of the Word “World” in M. Hoey, M. Mahlberg, M. Stubbs, T. Wolfgang (eds) *Text, Discourse, and Corpora*, London: Continuum, pp. 163-189.
- Tagliamonte, S. (2006) *Analysing Sociolinguistic Variation*, Cambridge: Cambridge University Press.
- Tarone, E., and Parrish, B. (1988) ‘Task-related variation in interlanguage: The case of articles’, *Language Learning* 38: 21-43.
- Thomas, M. (1989) ‘The acquisition of English articles by first- and second-language learners’, *Applied Psycholinguistics* 10: 335-355.
- Trenkic, D. (2007) ‘Variability in second language article production: beyond the representational deficit vs. processing constraints debate’, *Second Language Research* 23: 289 -327.
- Trenkic, D. (2004) ‘Definiteness in Serbian/Croatian/Bosnian and some implications for the general structure of the nominal phrase’, *Lingua* 114: 1401-1427.
- Trudgill, P. (2004) *New Dialect Formation: The Inevitability of Colonial Englishes*, Edinburgh: Edinburgh University Press.
- Vangsnes, O. (2001) ‘On noun phrase architecture, referentiality, and article systems’, *Studia Linguistica* 55: 249-299.
- Wahid, R. (2008) ‘The Use of Articles in American English, British English, Singapore English and Indian English: A Corpus-based Comparative Study’, *Paper presented at the first triennial Conference of the International Society for the Linguistics of English*, University of Freiburg, Freiburg. 8 – 11 October.

- Wahid, R. (2007) 'Capturing Variation in the Use of the Definite Article in Native and Nativised Varieties of English', *Paper presented at the annual Corpus Linguistics Conference*, University of Birmingham, Birmingham. 27 – 30 July.
- Wee, L. (2004) 'Singapore English: morphology and syntax', in B. Kortmann et al. (eds.) *A Handbook of Variation of English, Vol. 2: Morphology and Syntax* Berlin and New York: Mouton de Gruyter, pp. 1058-1072.
- Wee, L. and Ansaldo, U. (2004) 'Nouns and noun phrases', in L. Lim (ed) *Singapore English: A Grammatical Description*, Amsterdam and Philadelphia: John Benjamins, pp. 57-72.
- Wray, A. (2002) *Formulaic Language and the Lexicon*, Cambridge: Cambridge University Press.
- Young, R. (1996) 'Form-function relations in articles in English interlanguage', in R. Bayley and D. Preston (eds) *Second Language Acquisition and Linguistic Variation*, Amsterdam: John Benjamins, pp. 135-175.
- Zdorenko, T. and Paradis, J. (2008) 'The acquisition of articles in child second language English: Fluctuation, transfer or both?', *Second Language Research* 24: 227-250.

Appendix

Top 500 R-1 collocates of *the* in ICE-GB (unedited)

N	Word	N	Word	N	Word	N	Word	N	Word
1	FIRST	101	THINGS	201	NINETEEN	301	PRESENCE	401	TRAIN
2	SAME	102	WEST	202	RESULTS	302	REAR	402	TROUBLE
3	THE	103	NAME	203	ROLE	303	ROMANS	403	ABSENCE
4	OTHER	104	PROBLEMS	204	BOARD	304	SERVICE	404	AM
5	WAY	105	SOUTH	205	IRAQIS	305	SIXTH	405	CORRECT
6	TIME	106	CAR	206	LATTER	306	SUPPORT	406	DEGREE
7	UH	107	SUBJECT	207	LOWER	307	WEEKEND	407	EVENING
8	MOST	108	AMERICAN	208	MARKET	308	AMERICANS	408	FREE
9	END	109	FORM	209	TOTAL	309	CONFLICT	409	GIRL
10	WORLD	110	LONG	210	WRONG	310	FIVE	410	INSIDE
11	LAST	111	ORIGINAL	211	ARTS	311	FLOOR	411	LOSS
12	MOMENT	112	THREE	212	EARTH	312	GARDEN	412	MAP
13	WHOLE	113	FIELD	213	ENVIRONMENT	313	ORDER	413	MONARCHY
14	NEXT	114	SUN	214	KEY	314	POST	414	MOVEMENT
15	UNITED	115	DEPARTMENT	215	LIFE	315	RECORD	415	NET
16	ONE	116	AGE	216	NEAR	316	ROMAN	416	OIL
17	THIRD	117	BOOK	217	PERIOD	317	RULES	417	OPPOSITE
18	SECOND	118	BASIC	218	ALLIED	318	S	418	PICTURE
19	RIGHT	119	QUEEN	219	COMPUTER	319	SINGLE	419	SHORT
20	TWO	120	RESULT	220	CORNER	320	TAX	420	TWENTIETH
21	BALL	121	STATE	221	INITIAL	321	VARIOUS	421	WOMAN
22	BEST	122	WORD	222	OUTSIDE	322	VIEW	422	AFTERNOON
23	NEW	123	ACTUAL	223	PLAINTIFF	323	YEARS	423	BALANCE
24	CASE	124	MAN	224	QUALITY	324	ACTION	424	BLACK
25	FACT	125	MORNING	225	TAPE	325	AUDIENCE	425	CABINET
26	GULF	126	NEED	226	TRUTH	326	BANK	426	CONCEPT
27	MAIN	127	BIG	227	BRAIN	327	BIGGEST	427	CONTEXT
28	HOUSE	128	EFFECT	228	ECONOMY	328	BUSINESS	428	COW
29	SOVIET	129	HIGH	229	MEN	329	DETAILS	429	HEAT
30	BRITISH	130	MATTER	230	NERVE	330	FOURTH	430	HUMAN
31	WAR	131	WATER	231	NIGHT	331	GOVERNMENTS	431	JURY
32	GOVERNMENT	132	COURT	232	RATE	332	IMAGE	432	KNOWLEDGE
33	TOP	133	FOREIGN	233	ROOM	333	MILITARY	433	LANGUAGE
34	UHM	134	SITUATION	234	SEA	334	OPERATION	434	LARGE
35	MIDDLE	135	CHILDREN	235	WORDS	335	PRICE	435	LEAST
36	PEOPLE	136	HOME	236	DEFENCE	336	PROPOSED	436	MATERIAL
37	THIRD	137	CONSERVATIVE	237	I	337	TRADITIONAL	437	MIND
38	NATIONAL	138	DEVELOPMENT	238	MINISTER	338	VESSEL	438	MODERN
39	PAST	139	LATE	239	OTHERS	339	WINDOW	439	ON
40	MORE	140	UNIVERSITY	240	POSITION	340	CHAIRMAN	440	ONES
41	COUNTRY	141	WALL	241	REASONS	341	CHANCELLOR	441	PERFORMANCE
42	DAY	142	LAND	242	UK	342	COUNTRYSIDE	442	PRESIDENT
43	POINT	143	COST	243	AUTHOR	343	DISTRIBUTION	443	RANGE
44	BACK	144	FULL	244	EAST	344	HARD	444	SENATE
45	POLICE	145	REAL	245	EXTENT	345	HIGHEST	445	SEVENTH
46	PRIME	146	ATMOSPHERE	246	FORMER	346	INDUSTRY	446	V
47	PROBLEM	147	CENTRAL	247	INDIVIDUAL	347	LORD	447	VILLAGE
48	FUTURE	148	IRAQI	248	INTERNATIONAL	348	MUSIC	448	ARGUMENT
49	REST	149	START	249	UPPER	349	REFEREE	449	ASSOCIATION
50	NUMBER	150	ISSUE	250	YOUNG	350	STEPS	450	BATTLE
51	THING	151	LINE	251	ABILITY	351	STREET	451	CHILD
52	VERY	152	MAJOR	252	DEFENDANTS	352	TYPE	452	COMMON
53	OLD	153	PREVIOUS	253	MEDICAL	353	ABOVE	453	CONSEQUENCES
54	QUESTION	154	SENSE	254	METROPOLITAN	354	BLUE	454	EIGHTH
55	CENTRE	155	ACCIDENT	255	SOUND	355	FAMOUS	455	ENGLAND
56	COURSE	156	CHURCH	256	WHITE	356	GAME	456	EXCHANGE
57	IDEA	157	GROUP	257	DATA	357	IMPRESSION	457	EYES
58	FRONT	158	POPULATION	258	DECISION	358	LITTLE	458	GERMAN
59	WORK	159	POSSIBILITY	259	DIFFERENCE	359	MID	459	GREEN
60	LEFT	160	MONEY	260	E	360	PLAY	460	HISTORY
61	ROAD	161	NATURE	261	GOOD	361	PRESS	461	LACK
62	AREA	162	SOCIAL	262	GREATEST	362	PRIVATE	462	LARGER
63	PUBLIC	163	ALLIES	263	HEALTH	363	PROSPECT	463	NINETEENTH
64	EARLY	164	COLLEGE	264	INFORMATION	364	REGION	464	POOL
65	EUROPEAN	165	SUMMER	265	LEVEL	365	RIVER	465	POOR
66	SORT	166	POLL	266	NORMAL	366	SECURITY	466	PRIMARY
67	SYSTEM	167	WORST	267	OPPORTUNITY	367	SITE	467	PRODUCTION
68	YEAR	168	BASIC	268	PATIENT	368	TRIAL	468	RELATIONSHIP
69	PARTY	169	DOOR	269	RED	369	ACT	469	SHAPE
70	LAW	170	EVIDENCE	270	STRUCTURE	370	ARMY	470	TITLE
71	GREAT	171	FAMILY	271	TERM	371	CELL	471	AVERAGE
72	CURRENT	172	IN	272	USER	372	CROSS	472	BOOKS
73	LIGHT	173	LATEST	273	WEEK	373	DIRECTION	473	BORDER
74	FINAL	174	PHONE	274	MEETING	374	ECONOMIC	474	CAUSE
75	CITY	175	SIDE	275	POWER	375	EIGHTEENTH	475	CROWN
76	COMMUNITY	176	VEHICLE	276	PROJECT	376	ENGINE	476	HALFWAY
77	NORTH	177	B	277	SMALL	377	EXTRA	477	IMPORTANCE
78	GROUND	178	COUNCIL	278	STUDENTS	378	FIGURE	478	KITCHEN
79	HONOURABLE	179	EDGE	279	TASK	379	FIGURES	479	MAT
80	BEGINNING	180	FOUR	280	TEXT	380	LIBRARY	480	MOTOR
81	AMOUNT	181	MAJORITY	281	US	381	MEDIA	481	PROGRAMME
82	GENERAL	182	PLACE	282	C	382	PART	482	R
83	U	183	POLITICAL	283	DIFFERENT	383	RECENT	483	SCENE
84	SURFACE	184	SCHOOL	284	LEADER	384	RELEVANT	484	SIX
85	HEAD	185	SECRETARY	285	MASS	385	SCHEME	485	SOIL
86	PROCESS	186	WORLDS	286	POTENTIAL	386	STUDY	486	TELEPHONE
87	AIR	187	DATE	287	TABLE	387	TORIES	487	TELEVISION
88	LOCAL	188	DESIGN	288	BACKGROUND	388	USUAL	488	THEORY
89	USE	189	FIFTH	289	COLD	389	DISTANCE	489	TIMES
90	BOTTOM	190	JOB	290	FILM	390	EVENT	490	UNION
91	FRENCH	191	LEAD	291	LETTER	391	FACTS	491	WEATHER
92	COMPANY	192	STORY	292	PENALTY	392	IMPACT	492	ANCIENT
93	LABOUR	193	BUILDING	293	STANDARD	393	IRISH	493	BLOOD
94	PRESENT	194	ENGLISH	294	TERMS	394	MUSCLE	494	CHIEF
95	TWENTY	195	PERSON	295	WESTERN	395	OFFICIAL	495	CIRCUMSTANCES
96	KIND	196	SIZE	296	A	396	PHYSICAL	496	CROWD
97	ROYAL	197	ANSWER	297	EFFECTS	397	REGISTRATION	497	DEBATE
98	FOLLOWING	198	BODY	298	FACE	398	REPORT	498	DEVICE
99	REASON	199	LONDON	299	MODEL	399	ROOF	499	DOCTOR
100	FAR	200	NEWS	300	OPENING	400	THEATRE	500	ENERGY

Top 500 R-1 collocates of *the* in ICE-AUS (unedited)

N	Word	N	Word	N	Word	N	Word	N	Word
1	SAME	101	JOB	201	PREVIOUS	301	I	401	REGION
2	FIRST	102	SYDNEY	202	START	302	LANGUAGE	402	TOILET
3	OTHER	103	COUNTRY	203	TEXT	303	LETTER	403	VIEW
4	THE	104	PROCESS	204	VARIOUS	304	NRMA	404	WIND
5	WAY	105	SOUTH	205	FIELD	305	RELEVANT	405	ACCUSED
6	TIME	106	BALL	206	GREEN	306	SENSE	406	APPEACE
7	MOST	107	BOARD	207	RESEARCH	307	TERM	407	ART
8	LAST	108	CHILD	208	STATES	308	TRADITIONAL	408	ASSOCIATION
9	END	109	PHONE	209	STORY	309	BUDGET	409	BALANCE
10	WORLD	110	PLACE	210	WEST	310	COMPUTER	410	CLASS
11	AUSTRALIAN	111	SYSTEM	211	BEACH	311	CONSTITUTION	411	COLD
12	NEXT	112	BEGINNING	212	CENTRAL	312	CROWD	412	CORRECT
13	ONE	113	CURRENT	213	GAME	313	EFFECT	413	ER
14	WHOLE	114	HIGH	214	HISTORY	314	INDUSTRIAL	414	GALLERY
15	NEW	115	LAW	215	MAJORITY	315	INDUSTRY	415	HOLIDAYS
16	GOVERNMENT	116	MONEY	216	WORDS	316	JAPANESE	416	IMPORTANCE
17	ONLY	117	PRESENT	217	FLOOR	317	PROJECT	417	ISLAND
18	AH	118	THING	218	QUALITY	318	REASONS	418	KITCHEN
19	BEST	119	COMMUNITY	219	WOMEN	319	RELATIONSHIP	419	MAXIMUM
20	UM	120	COMMONWEALTH	220	ACT	320	ROOM	420	PROPOSED
21	TWO	121	ENVIRONMENT	221	FIRE	321	WEATHER	421	RED
22	RIGHT	122	NAME	222	FOOD	322	WORKPLACE	422	SITE
23	FACT	123	NEED	223	KIND	323	CONTRACT	423	SUMMER
24	MOMENT	124	FINAL	224	MEMBERS	324	DIFFERENT	424	U
25	SECOND	125	IN	225	NSW	325	GREATEST	425	WINDOW
26	TOP	126	MAJOR	226	POLITICAL	326	HIGHEST	426	APPROPRIATE
27	PEOPLE	127	DEVELOPMENT	227	RANGE	327	LEAD	427	BASE
28	DAY	128	KEY	228	WOMAN	328	MEN	428	CAUSE
29	STATE	129	MAN	229	BIGGEST	329	PERIOD	429	COMMON
30	CASE	130	DIFFERENCE	230	CHAR	330	POLICY	430	CONCLUSION
31	CAR	131	LONG	231	CHARGE	331	PROGRAM	431	DÉGRÉE
32	MAIN	132	REASON	232	CONCEPT	332	PURPOSE	432	DRIVE
33	POLICE	133	SIA	233	ECONOMIC	333	WALL	433	EFFECTS
34	YEAR	134	TWENTY	234	FORM	334	BLUE	434	ELECTION
35	PROBLEM	135	CHILDREN	235	TYPE	335	BUS	435	EXISTING
36	HOUSE	136	MORNING	236	UNION	336	CORRIDOR	436	FRENCH
37	POINT	137	POWER	237	VEHICLE	337	GOVERNOR	437	GIRLS
38	EVIDENCE	138	COAST	238	WEEK	338	GROWTH	438	GRAND
39	FEDERAL	139	SIDE	239	WORLDS	339	LACK	439	GUY
40	UNIVERSITY	140	SITUATION	240	BLACK	340	LIVER	440	INCIDENT
41	NUMBER	141	A	241	COMMITTEE	341	OUTSIDE	441	LARGER
42	PAST	142	FORMER	242	FOUR	342	TRADE	442	LEAST
43	FAMILY	143	POSITION	243	SIB	343	USUAL	443	LEGISLATION
44	QUESTION	144	DOOR	244	SHORT	344	CHANCE	444	LOW
45	FRONT	145	ROLE	245	SOVIET	345	CONTEXT	445	MOTHER
46	BACK	146	LEFT	246	STREET	346	DATA	446	PREMIER
47	UNITED	147	MEETING	247	AMERICAN	347	EVENT	447	PRIMARY
48	VERY	148	PARTY	248	CHANGES	348	HEALTH	448	PRODUCT
49	PUBLIC	149	PROBLEMS	249	COALITION	349	LIBRARY	449	PROPOSAL
50	UH	150	STUDY	250	DARK	350	NET	450	RECENT
51	NATIONAL	151	TABLE	251	ENTIRE	351	OVERALL	451	ROOF
52	OLD	152	AGE	252	NORTHERN	352	PRESSURE	452	SPEED
53	ROAD	153	BUSINESS	253	WAR	353	QUEEN	453	TRAIN
54	PRIME	154	DECISION	254	ABILITY	354	REF	454	TREASURER
55	WATER	155	NATURE	255	ABORIGINAL	355	SEA	455	AFTERNOON
56	BIG	156	OFFICE	256	BANK	356	SERVICE	456	BAR
57	GREAT	157	ONES	257	GOOD	357	SKIN	457	BRIDGE
58	PERSON	158	FILM	258	WEEKEND	358	SOIL	458	CIRCUMSTANCES
59	REST	159	GROUP	259	WRONG	359	STAGE	459	CLUB
60	GROUND	160	HUMAN	260	ABOVE	360	STATE'S	460	CONDITIONS
61	MIDDLE	161	INFORMATION	261	COMMISSION	361	TOWN	461	DISTANCE
62	WORD	162	OPPOSITION	262	DATE	362	VICTORIAN	462	DOGS
63	CITY	163	SIZE	263	INTERNATIONAL	363	ANNUAL	463	DOMINANT
64	DEFENDANT	164	SUBJECT	264	JURY	364	BOAT	464	EAST
65	LITTLE	165	FULL	265	PLANT	365	DRIVER	465	ENGINE
66	MATTER	166	LATEST	266	PRESS	366	FINANCIAL	466	FEW
67	US	167	LIGHT	267	RATE	367	LOWER	467	GYM
68	FOLLOWING	168	PAPER	268	RESULTS	368	MATCH	468	HOSPITAL
69	MORE	169	ACTUAL	269	SOUTHERN	369	NON	469	LABOUR
70	THREE	170	GOVERNMENTS	270	SURFACE	370	ON	470	LIGHTS
71	EARLY	171	HOME	271	WHITE	371	PICTURE	471	NATURAL
72	MARKET	172	SUN	272	EXTENT	372	SMALL	472	PERFORMANCE
73	NIGHT	173	TAX	273	FACTS	373	STAFF	473	POST
74	WORK	174	WORST	274	GOLD	374	STANDARD	474	QUEENSLAND
75	BOTTOM	175	ECONOMY	275	HARBOUR	375	TERMS	475	SUGAR
76	LAND	176	LABOR	276	MEMBER	376	WESTERN	476	TAPE
77	SORT	177	LIBERAL	277	NEWS	377	YEARS	477	TEST
78	USE	178	AMOUNT	278	PROSECUTION	378	ABC	478	TIMBER
79	DEPARTMENT	179	CENTRE	279	RULES	379	DROUGHT	479	TROPICAL
80	BASIS	180	GENERAL	280	TRUTH	380	FACE	480	WILD
81	COUNCIL	181	LINE	281	YOUNG	381	INTRODUCTION	481	ALL
82	THIRD	182	NINETEEN	282	BUILDING	382	LEAGUE	482	BACKGROUND
83	AREA	183	OPPORTUNITY	283	EVENING	383	MID	483	CAMERA
84	IDEA	184	POPULATION	284	HEART	384	MUSIC	484	CONSUMER
85	ISSUE	185	RACE	285	IMPACT	385	PARTIES	485	DEFENCE
86	KIDS	186	REPORT	286	INDIVIDUAL	386	SPOT	486	DOCTOR
87	MEDIA	187	RIVER	287	INSIDE	387	STREETS	487	EDGE
88	REAL	188	SOCIAL	288	LARGEST	388	TREE	488	FUNCTION
89	SCHOOL	189	AIR	289	LATTER	389	TRIP	489	INTERVIEW
90	BOOK	190	CROWN	290	MATERIAL	390	WORKFORCE	490	LEADER
91	HEAD	191	LATE	291	PART	391	ANSWER	491	MALE
92	ORIGINAL	192	MINISTER	292	POTENTIAL	392	APPLICATION	492	NATION
93	RESULT	193	NORTH	293	STUDENTS	393	BASIC	493	OPEN
94	THINGS	194	PRICE	294	TRACK	394	BOYS	494	PAGE
95	CORNER	195	ROYAL	295	VALUE	395	CHAIRPERSON	495	POSSIBILITY
96	COURT	196	TOTAL	296	YOU	396	CHIEF	496	PRESENCE
97	COMPANY	197	BODY	297	AVERAGE	397	FENCE	497	RESIDENT
98	COURSE	198	LEVEL	298	BRITISH	398	HIGHER	498	QUESTIONS
99	COST	199	LOCAL	299	BUSH	399	ORGANISATION	499	SAND
100	FUTURE	200	OTHERS	300	FIVE	400	READER	500	SEAT

Top 500 R-1 collocates of *the* in ICE-NZ (unedited)

N	Word	N	Word	N	Word	N	Word	N	Word
1	FIRST	101	ALL	201	TRIBUNAL	301	POTENTIAL	401	M
2	SAME	102	HEALTH	202	AGE	302	QUESTIONS	402	SCENE
3	OTHER	103	MAJOR	203	BLUE	303	SERVICE	403	TEAM
4	NEW	104	SEA	204	BUDGET	304	SOUND	404	UPPER
5	THE	105	MAN	205	CENTRE	305	TABLE	405	VALUE
6	TIME	106	BEGINNING	206	DATA	306	WOMAN	406	WELFARE
7	WAY	107	CHILDREN	207	EMPLOYMENT	307	A	407	WESTERN
8	LAST	108	WELLINGTON	208	LARGE	308	EFFECT	408	DOMINANT
9	END	109	BANK	209	WHITE	309	LARGEST	409	EVENT
10	GOVERNMENT	110	LAW	210	WOMEN	310	ON	410	FILM
11	MOST	111	HIGH	211	WORST	311	PACIFIC	411	FOOD
12	NEXT	112	BOAT	212	CORNER	312	PARENTS	412	GROWTH
13	ER	113	ROOM	213	ELECTION	313	PERIOD	413	LEAGUE
14	ONLY	114	DOOR	214	EMPLOYEE	314	REASONS	414	MARAE
15	WORLD	115	GROUP	215	REASON	315	STAGE	415	MOTORWAY
16	TWO	116	STORY	216	WAR	316	ABOVE	416	NORTHERN
17	SECOND	117	TOTAL	217	FORM	317	AMERICAN	417	ACTION
18	BALL	118	GAME	218	FORMER	318	BIGGEST	418	CLASS
19	ONE	119	LOCAL	219	GUY	319	BUSINESS	419	COMPANY
20	MOMENT	120	ECONOMY	220	SOCIAL	320	CENTRAL	420	EDGE
21	WHOLE	121	SCHOOL	221	BASE	321	COMPUTER	421	GREATEST
22	HOUSE	122	COST	222	GOOD	322	EARTH	422	HILL
23	PEOPLE	123	INFORMATION	223	LATTER	323	ENGLISH	423	HISTORY
24	BEST	124	LAND	224	LEVEL	324	ENTIRE	424	INDIVIDUAL
25	FACT	125	PROCESS	225	QUALITY	325	FIVE	425	LOUNGE
26	NATIONAL	126	RIVER	226	U	326	MEDIA	426	NATURAL
27	RIGHT	127	REAL	227	WALL	327	OUTSIDE	427	ORGANISATION
28	LM	128	BOARD	228	DECISION	328	OVERALL	428	PERSONAL
29	CROWN	129	LITTLE	229	FIRE	329	PERFORMANCE	429	POEM
30	MAIN	130	SHIP	230	KIWI	330	PLANT	430	PROGRAM
31	POLICE	131	BRITISH	231	LIBRARY	331	PRESENCE	431	PUB
32	TOP	132	GOVERNMENTS	232	POPULATION	332	ROYAL	432	REGION
33	BACK	133	LATE	233	BEACH	333	STANDARD	433	ROOT
34	PUBLIC	134	NEWS	234	BENEFIT	334	TELEVISION	434	SHORE
35	OLD	135	SOIL	235	BODY	335	WITNESS	435	SKY
36	REST	136	YEARS	236	EFFECTS	336	BABY	436	START
37	COUNTRY	137	DEVELOPMENT	237	ISLAND	337	EVENING	437	TAPE
38	MAORI	138	JOB	238	PAPER	338	HUMAN	438	TEXT
39	SOUTH	139	AMOUNT	239	PRIVATE	339	I	439	TRADITIONAL
40	MEMBER	140	NINETEEN	240	PROGRAMME	340	KIWIS	440	WORKPLACE
41	PAST	141	GENERAL	241	ABILITY	341	MILITARY	441	C
42	PROBLEM	142	PREVIOUS	242	ENVIRONMENT	342	MOTHER	442	COST
43	NUMBER	143	UNIVERSITY	243	FOURTH	343	MUSIC	443	DOCTOR
44	DAY	144	IDEA	244	OFFICE	344	ORDER	444	GARDEN
45	MIDDLE	145	PHONE	245	OTHERS	345	POSITION	445	IMPACT
46	WORK	146	PLACE	246	POLITICAL	346	PRICE	446	MOUNTAIN
47	CASE	147	SITUATION	247	PROJECT	347	RANGE	447	PASTURE
48	CAR	148	SIZE	248	RESULT	348	TERM	448	RACE
49	THING	149	YOUNG	249	WINDOW	349	AFTERNOON	449	SENSE
50	POINT	150	BASIS	250	ECONOMIC	350	BASIC	450	STATES
51	GROUND	151	PRIME	251	EXTENT	351	BIBLE	451	STATION
52	MINISTER	152	SUN	252	KIND	352	COMMITTEE	452	SUPPORT
53	NORTH	153	AVERAGE	253	MATTER	353	LOSS	453	WINDOWS
54	FUTURE	154	CHURCH	254	NATURE	354	PRESSURE	454	APPLICATION
55	FOLLOWING	155	COURT	255	OPPOSITION	355	STRUCTURE	455	CHARACTER
56	WATER	156	CURRENT	256	PLAY	356	EMPLOYER	456	CHARGE
57	NIGHT	157	LONG	257	S	357	FACE	457	CONFERENCE
58	MORE	158	POWER	258	STUDY	358	GIRLS	458	CROWD
59	PERSON	159	RESULTS	259	WORLDS	359	HARBOUR	459	CUP
60	SYSTEM	160	AIR	260	BAR	360	PROPERTY	460	DEATH
61	WORD	161	BILL	261	CLYDE	361	PROPOSED	461	EYE
62	YEAR	162	FRENCH	262	DEFENCE	362	TRUST	462	LORD
63	FAMILY	163	KIDS	263	KEY	363	VIEW	463	MESSAGE
64	FRONT	164	ONES	264	LETTER	364	WALLS	464	NUMBERS
65	QUESTION	165	FLOOR	265	POOR	365	WORDS	465	PART
66	EARLY	166	HEAD	266	RISK	366	BUS	466	POLICY
67	USE	167	MEN	267	SOUTHERN	367	CHANGES	467	RED
68	CITY	168	ORIGINAL	268	SURFACE	368	COSTS	468	SPEECH
69	ROAD	169	PARTY	269	SURVEY	369	DEGREE	469	STAFF
70	BIG	170	CLUB	270	WEEKEND	370	DESIGN	470	BEDROOM
71	NAME	171	MARKET	271	WORKING	371	EAST	471	CAUSE
72	FIELD	172	MINISTRY	272	FISH	372	EVENTS	472	CHANGE
73	CHILD	173	AUDIENCE	273	HIGHEST	373	FOREST	473	EARLIER
74	COURSE	174	COUNCIL	274	ISSUES	374	LIST	474	EYES
75	UNITED	175	GREAT	275	KITCHEN	375	POSSIBILITY	475	FACTS
76	ACCUSED	176	ACTUAL	276	LANGUAGE	376	T	476	FAR
77	LABOUR	177	DIFFERENCE	277	MEETING	377	WEATHER	477	FIGURES
78	NEED	178	LEFT	278	STUDENT	378	BACKHAND	478	FRONTMAN
79	SORT	179	SOVIET	279	UNION	379	CONCEPT	479	HAZARD
80	TWENTY	180	MAJORITY	280	DOG	380	DARK	480	IMPORTANCE
81	COMPANY	181	AUSTRALIANS	281	HOME	381	HOTEL	481	LEAST
82	PRESENT	182	BAY	282	HONOURABLE	382	MEAT	482	LENGTH
83	MORNING	183	LATEST	283	INTERNATIONAL	383	OPEN	483	MATERIAL
84	THIRD	184	PURPOSE	284	JURY	384	OWNER	484	NATIVE
85	EVIDENCE	185	STREET	285	REPORT	385	QUESTIONNAIRE	485	RULES
86	AREA	186	BOOK	286	SMALL	386	RATE	486	TEST
87	AUCKLAND	187	FOUR	287	TYPE	387	SALE	487	TRUTH
88	THREE	188	IN	288	LOW	388	TEN	488	USUAL
89	LINE	189	PROBLEMS	289	NET	389	1980S	489	WINTER
90	BOTTOM	190	WEST	290	ROLE	390	5	490	ARTS
91	COMMUNITY	191	ACT	291	TASK	391	BACKGROUND	491	BOYS
92	TREATY	192	LIGHT	292	TREE	392	CAMP	492	CONCENTRATION
93	FINAL	193	LOWER	293	VARIOUS	393	CHANGE	493	DEAL
94	MONEY	194	OPPORTUNITY	294	WRONG	394	CHIEF	494	DEBATE
95	VERY	195	SIDE	295	ANSWER	395	DIFFERENT	495	DISTRICT
96	DEPARTMENT	196	WEEK	296	B	396	EXISTING	496	FARM
97	AUSTRALIAN	197	WIND	297	BENEFITS	397	FENCE	497	FEW
98	ISSUE	198	BUILDING	298	FULL	398	FOREHAND	498	GREATER
99	STATE	199	GREEN	299	MEAN	399	INTRODUCTION	499	HARD
100	THINGS	200	SUBJECT	300	POST	400	LEADER	500	LIFE

Top 500 R-1 collocates of *the* in the US data (unedited)

N	Word	N	Word	N	Word	N	Word	N	Word
1	FIRST	101	RESULT	201	MUSIC	301	A04	401	J17
2	SAME	102	ROOM	202	PICTURE	302	A20	402	J41
3	MOST	103	SUN	203	POLKA	303	A22	403	J54
4	OTHER	104	COMPANY	204	TABLE	304	A44	404	J57
5	UNITED	105	COST	205	TRUTH	305	ABSENCE	405	J63
6	U	106	DATA	206	VARIOUS	306	B16	406	J69
7	NEW	107	FINAL	207	WALL	307	BIGGEST	407	K04
8	WORLD	108	GROUND	208	AGE	308	BOOK	408	KIND
9	TIME	109	POLITICAL	209	BOTTOM	309	BRAIN	409	LARGE
10	ONLY	110	BUSH	210	BUSINESS	310	E10	410	LEFT
11	SECOND	111	GROUP	211	DIFFERENCE	311	E16	411	LIGHT
12	NEXT	112	LINE	212	DISEASE	312	E17	412	MEAN
13	WAY	113	MARKET	213	EDITOR	313	E25	413	OPPOSITE
14	TWO	114	SOVIET	214	INDIVIDUAL	314	E29	414	PHONE
15	LAST	115	SURFACE	215	J36	315	E30	415	SMALL
16	BEST	116	BOARD	216	J58	316	E36	416	STANDARD
17	AMERICAN	117	EIGN	217	J62	317	EXISTENCE	417	TRADITIONAL
18	OLD	118	ENTIRE	218	MATERIAL	318	EXTENT	418	VACCINE
19	WHITE	119	HUMAN	219	MILITARY	319	F14	419	YEARS
20	END	120	J06	220	NATURAL	320	F35	420	A01
21	STATE	121	J75	221	NEWS	321	HEAT	421	A27
22	TOP	122	LONG	222	POSSIBILITY	322	INDUSTRY	422	A40
23	HOUSE	123	RESULTS	223	PREVIOUS	323	INTERNATIONAL	423	B09
24	PAST	124	STUDY	224	SAT	324	J12	424	BANK
25	PRESIDENT	125	SUBJECT	225	SENSE	325	J14	425	BRIDGE
26	CASE	126	AVERAGE	226	WORST	326	J40	426	CAR
27	CITY	127	COURSE	227	B14	327	J66	427	DEATH
28	GENERAL	128	DEVELOPMENT	228	COMMON	328	K10	428	E04
29	NATIONAL	129	LOWER	229	COMPANYS	329	MORNING	429	E14
30	WAR	130	PROGRAM	230	COURT	330	OPENING	430	E21
31	PUBLIC	131	SOUTH	231	DREAM	331	PATENTING	431	ENGLISH
32	RIGHT	132	TOTAL	232	F17	332	PROBLEMS	432	EVIDENCE
33	MAJOR	133	WATER	233	F33	333	RACE	433	F01
34	ONE	134	AIR	234	FIELD	334	RELATIVE	434	F12
35	GAME	135	ECONOMY	235	FLOW	335	SIDE	435	F41
36	WHOLE	136	EFFECT	236	J04	336	STOCK	436	F44
37	EARLY	137	J22	237	LOWEST	337	SUPREME	437	FREE
38	NUMBER	138	J23	238	MEN	338	WORDS	438	HISTORY
39	MAIN	139	LAND	239	MORAL	339	1980S	439	IMAGE
40	GOVERNMENT	140	LOCAL	240	PARTY	340	A12	440	INITIAL
41	COUNTRY	141	MODEL	241	PERSON	341	A14	441	IRA
42	PRESS	142	NATION	242	RABBI	342	ANSWER	442	IRISH
43	REST	143	NEED	243	RATE	343	B04	443	J08
44	BIG	144	WORD	244	REAL	344	B13	444	J27
45	E	145	WORLDS	245	SELF	345	CAMPAIGN	445	J55
46	FACT	146	AREA	246	SENATE	346	DARK	446	J77
47	PROBLEM	147	HIGHEST	247	SOCIAL	347	DECISION	447	K16
48	DAY	148	IMPORTANCE	248	TEAM	348	DNA	448	MACHINE
49	SYSTEM	149	KEY	249	TREATY	349	E07	449	MARK
50	USE	150	NIGHT	250	A39	350	E08	450	MATTER
51	VERY	151	BASIS	251	CAMERA	351	E19	451	PRESIDENTIAL
52	DOOR	152	BODY	252	CITYS	352	EDGE	452	REASON
53	FOLLOWING	153	ECONOMIC	253	DEGREE	353	EXAMINER	453	RULES
54	FRONT	154	EFFECTS	254	E13	354	F13	454	SIGN
55	MAN	155	J02	255	E35	355	FIBER	455	SOUTHERN
56	MIDDLE	156	J31	256	F19	356	FULL	456	STREETS
57	MORE	157	PROCESS	257	F22	357	GOAL	457	STRUCTURE
58	ISSUE	158	STORY	258	F24	358	HOME	458	TREE
59	THREE	159	CENTRAL	259	F32	359	J01	459	VICTIM
60	FEDERAL	160	CONCEPT	260	FACE	360	J11	460	WELL
61	QUESTION	161	E03	261	FALL	361	J20	461	1
62	ROLE	162	EC	262	INFORMATION	362	J35	462	A06
63	LAW	163	F20	263	ISLAND	363	J50	463	A19
64	THIRD	164	F28	264	J07	364	J52	464	A26
65	BRITISH	165	F43	265	J19	365	J79	465	A33
66	GREAT	166	FOUR	266	J24	366	K07	466	B06
67	IDEA	167	GREATEST	267	J56	367	K08	467	B08
68	PEOPLE	168	J67	268	JUDGE	368	MIND	468	B20
69	POWER	169	ORIGINAL	269	KITCHEN	369	PART	469	B24
70	WORK	170	PROJECT	270	LARGEST	370	PLACE	470	CELLS
71	BILL	171	SCHOOL	271	LATEST	371	PRIMARY	471	CORNER
72	POINT	172	SITUATION	272	NAME	372	RADIO	472	COUNTRY'S
73	BLACK	173	SIZE	273	NATURE	373	SOUND	473	DIRECTION
74	CURRENT	174	UPPER	274	NOTION	374	SUMMER	474	ENVIRONMENT
75	JOB	175	BUILDING	275	OBSERVED	375	TOWNS	475	EXACT
76	FAMILY	176	CHILDREN	276	POOR	376	VALUE	476	F03
77	LATE	177	COMMUNITY	277	POPULATION	377	WOMEN	477	F15
78	LATTER	178	CONTEXT	278	PROBABILITY	378	X	478	F23
79	WOMAN	179	F21	279	REPUBLICAN	379	A11	479	F27
80	BASIC	180	F47	280	SHOW	380	A38	480	F37
81	EARTH	181	HOUSEHOLD	281	CIVIL	381	ABOVE	481	F40
82	FORM	182	J39	282	DEMOCRATIC	382	ADMINISTRATION	482	F46
83	J64	183	J44	283	E27	383	AMERICANS	483	INTRODUCTION
84	NATIONS	184	J72	284	F48	384	B27	484	ISSUES
85	STREET	185	J80	285	FLOOR	385	CHANGE	485	J15
86	WEST	186	MAJORITY	286	FRENCH	386	CHILD	486	J25
87	BACK	187	MOMENT	287	GOOD	387	CLUSTER	487	JOINT
88	BEGINNING	188	TEMPERATURE	288	HEART	388	E01	488	K02
89	FORMER	189	ABILITY	289	J03	389	E18	489	K29
90	FUTURE	190	ART	290	J26	390	ELECTION	490	LEVEL
91	HIGH	191	DESIGN	291	J51	391	F06	491	MANY
92	PRESIDENTS	192	F29	292	J53	392	F38	492	MEDIA
93	CENTER	193	IPA	293	J74	393	F42	493	PACIFIC
94	PRESENCE	194	J05	294	LEGAL	394	F45	494	POET
95	PRESENT	195	J16	295	MAP	395	FEW	495	POSITION
96	ROAD	196	J78	296	MID	396	FILM	496	QUALITY
97	UNIVERSITY	197	JAPANESE	297	POTENTIAL	397	FINALE	497	RIGHTS
98	YEAR	198	LEAST	298	PRESSURE	398	FOURTH	498	RULE
99	COLD	199	MODERN	299	SET	399	FUNDAMENTAL	499	SIGHT
100	NORTH	200	MONEY	300	TERM	400	J09	500	SKY

Top 500 R-1 collocates of *the* in ICE-SIN (unedited)

N	Word	N	Word	N	Word	N	Word	N	Word
1	FIRST	101	LONG	201	FULL	301	LIBRARY	401	KIDS
2	SAME	102	SEA	202	MALAYS	302	PICK	402	MATTER
3	OTHER	103	STRAITS	203	PLANTS	303	RED	403	PLAINTIFFS
4	THE	104	COURSE	204	PRODUCT	304	CUPBOARD	404	RECENT
5	SINGAPORE	105	LANGUAGE	205	SOUND	305	EXAMINATION	405	YOUNGER
6	WAY	106	LAW	206	UN	306	FACTS	406	AGED
7	GOVERNMENT	107	MALAY	207	HUMAN	307	HOME	407	BANK
8	NEXT	108	PRIME	208	LINE	308	INITIAL	408	BOARD
9	TIME	109	TOTAL	209	OPEN	309	LIGHT	409	HARD
10	LAST	110	BACK	210	PERIOD	310	NATION	410	HOSPITAL
11	WHOLE	111	IDEA	211	PREVIOUS	311	NORTH	411	INTRODUCTION
12	WORLD	112	DIFFERENT	212	ABILITY	312	PORT	412	LINGUISTIC
13	END	113	CURRENT	213	BASIC	313	POST	413	PEACE
14	NEW	114	KEY	214	BASIS	314	TEACHER	414	SOCIETY
15	UH	115	AIR	215	LATTER	315	THAI	415	SUM
16	SECOND	116	ECONOMIC	216	PROGRAMME	316	TRADE	416	TRUTH
17	TWO	117	SITUATION	217	AGE	317	AVERAGE	417	USER
18	WORD	118	WORK	218	CONSTRUCTION	318	BOYS	418	WEATHER
19	PEOPLE	119	NAME	219	FIVE	319	CONTEXT	419	WHAT
20	MOST	120	ROAD	220	OFFICE	320	FACILITY	420	WOMAN
21	ONE	121	DEPARTMENT	221	OPPORTUNITY	321	GOAL	421	BACKGROUND
22	CHINESE	122	DIFFERENCE	222	PAPER	322	HISTORY	422	BUS
23	MAIN	123	PART	223	SCHEME	323	LATE	423	CHANGE
24	BALL	124	SENSE	224	AUDIENCE	324	LEVEL	424	CONVERSATION
25	NATIONAL	125	ANSWER	225	BETTER	325	MOTHER	425	DATABASE
26	PAP	126	LATEST	226	FOURTH	326	NATURAL	426	ESTABLISHMENT
27	BEST	127	TRADITIONAL	227	LETTER	327	POOR	427	EXISTING
28	ONLY	128	ECONOMY	228	LOWER	328	RECORD	428	FIGURE
29	US	129	GROUP	229	MILITARY	329	SCORE	429	HIGHEST
30	FACT	130	PLACE	230	MUSIC	330	SERVICE	430	JUDGE
31	PROBLEM	131	CENTRAL	231	ORIGINAL	331	SIX	431	LABOUR
32	PAST	132	COMMUNITY	232	PEOPLE'S	332	TEST	432	LARGE
33	CASE	133	JOB	233	POWER	333	WHITE	433	LITTLE
34	NUMBER	134	TWENTY	234	SHOW	334	WORKERS	434	MESSAGE
35	RIGHT	135	BOTTOM	235	TERM	335	ACTION	435	NET
36	UNIVERSITY	136	HEAD	236	THINGS	336	BLACK	436	NUS
37	REST	137	PLAINTIFF	237	ACCIDENT	337	BLUE	437	PRE
38	MARKET	138	PRICE	238	BUILDING	338	COUNTRY'S	438	REASONS
39	QUESTION	139	SOVIET	239	CAR	339	FILM	439	REMAINING
40	MOMENT	140	ENGLISH	240	COMMITTEE	340	FORTY	440	RISK
41	SYSTEM	141	REEF	241	GREAT	341	GREEN	441	SCREEN
42	FOLLOWING	142	RESULT	242	IMPORTANCE	342	GROWTH	442	SKIN
43	MORE	143	COST	243	PLANT	343	ISSUES	443	SOUTH
44	UNITED	144	INTERNATIONAL	244	SOIL	344	LEAD	444	STRUCTURAL
45	MINISTRY	145	ACTUAL	245	STAFF	345	OLDER	445	TEACHING
46	USE	146	DEVELOPMENT	246	WORLDS	346	PRODUCTION	446	TRAFFIC
47	FAMILY	147	RESULTS	247	CONTRACT	347	ROCK	447	VIETNAMESE
48	ARTS	148	ROLE	248	DATE	348	SIZE	448	VIEW
49	PUBLIC	149	BRITISH	249	FRONT	349	STANDARD	449	AFTERNOON
50	THREE	150	FORM	250	HDB	350	STUDENT	450	ANCIENT
51	TOP	151	PATIENT	251	INDONESIANS	351	SUN	451	CONTENT
52	COMPANY	152	YEARS	252	OPENING	352	TITLE	452	EXISTENCE
53	OLD	153	APPLICATION	253	STUDY	353	UPPER	453	FACE
54	THING	154	FOOD	254	ULTIMATE	354	WINDOW	454	FLAT
55	ABOVE	155	STATE	255	WEST	355	CIVIL	455	GIRL
56	DAY	156	YOUNG	256	YOU	356	CLIENT	456	GUY
57	OPPOSITION	157	EMPLOYER	257	CHIEF	357	DISABLED	457	HUSBAND
58	COURT	158	LOCAL	258	EMPLOYEE	358	ELDERLY	458	LIQUID
59	THIRD	159	MAJOR	259	EVENT	359	IN	459	LTA
60	POINT	160	NATURE	260	I	360	MEDICAL	460	MANY
61	COUNTRY	161	POLITICAL	261	NORMAL	361	MEETING	461	OWNER
62	ISSUE	162	DOOR	262	ONES	362	NECESSARY	462	P
63	MIDDLE	163	PURPOSE	263	PARENTS	363	NEEDS	463	PRESS
64	YEAR	164	DEFENCE	264	POLICY	364	NEWSPAPER	464	REPUBLIC
65	POPULATION	165	ENTIRE	265	PRIMARY	365	NINETEEN	465	SAID
66	CHILD	166	ISLAND	266	REAL	366	PAPERS	466	SDP
67	VARIOUS	167	LEFT	267	REFEREE	367	POTENTIAL	467	SERIES
68	STUDENTS	168	MORNING	268	REPORT	368	PRESENCE	468	SHOP
69	GENERAL	169	SUBJECT	269	SURFACE	369	RATE	469	SUPPORT
70	HOUSE	170	WORDS	270	USUAL	370	RULES	470	TEAM
71	POLICE	171	AREA	271	AMERICAN	371	START	471	WELL
72	PRESENT	172	BEGINNING	272	ART	372	TASK	472	WOMEN
73	REASON	173	DEFENDANT	273	CONFERENCE	373	WIFE	473	WRONG
74	MINISTER	174	FIELD	274	DATA	374	WORKING	474	APPELLANTS
75	SCHOOL	175	QUALITY	275	LIST	375	A	475	ATTENTION
76	REGION	176	SOCIAL	276	MEANING	376	DETAILS	476	BOOKS
77	BOOK	177	BUSINESS	277	OPERATION	377	ELECTION	477	BRAIN
78	EARLY	178	DECISION	278	PRESIDENT	378	FIFTH	478	COLOUR
79	KIND	179	DEFENDANTS	279	PROPOSED	379	HILL	479	CUSTOMER
80	MONEY	180	GOOD	280	RELATIONSHIP	380	LACK	480	EARLIER
81	PERSON	181	MOULD	281	TABLE	381	MOVE	481	EFFECT
82	COMPUTER	182	PHONE	282	BODY	382	OUTCOME	482	FIELDS
83	PROCESS	183	PROBLEMS	283	CITY	383	PERFORMANCE	483	INCREASING
84	NEWS	184	WATER	284	CLASS	384	SMALL	484	LADY
85	FUTURE	185	JAPANESE	285	COMMON	385	STATES	485	LEARNED
86	MAN	186	NIGHT	286	DOCTOR	386	WESTERN	486	LEAVES
87	CENTRE	187	PROPERTY	287	EAST	387	ASEAN	487	LIFE
88	INFORMATION	188	UHM	288	EDUCATION	388	ASIA	488	MANAGEMENT
89	STORY	189	PARTY	289	GOVERNMENTS	389	BULK	489	NARRATOR
90	VERY	190	ROOM	290	MEN	390	C	490	OVERALL
91	ENVIRONMENT	191	TYPE	291	OTHERS	391	CARD	491	PROGRAM
92	GROUND	192	VALUE	292	SAF	392	CHALLENGE	492	RAIN
93	NEED	193	BIG	293	SHARES	393	CONCEPTUAL	493	SOLUTION
94	CHILDREN	194	CONCEPT	294	STRUCTURE	394	CORRECT	494	STAGE
95	FINAL	195	EVIDENCE	295	TEACHERS	395	DEMAND	495	TEXT
96	PHILIPPINES	196	FORMER	296	FRENCH	396	DESIGN	496	WAR
97	FOUR	197	GAME	297	INDIAN	397	EXACT	497	WINDOWS
98	HIGH	198	MAJORITY	298	INDIVIDUAL	398	EYE	498	AIRPORT
99	PROJECT	199	PLAN	299	LAND	399	FOREIGN	499	ASIAN
100	AMOUNT	200	PRIVATE	300	LARGEST	400	IMPORTANT	500	BED

Top 500 R-1 collocates of *the* in ICE-IND (unedited)

N	Word	N	Word	N	Word	N	Word	N	Word
1	FIRST	101	PERSON	201	COST	301	FILM	401	EGG
2	SAME	102	SYSTEM	202	INCIDENT	302	FOREIGN	402	GAS
3	OTHER	103	MORNING	203	LEARNED	303	NET	403	HEALTH
4	PEOPLE	104	REASON	204	MINISTER	304	OFFICE	404	INFLUENCE
5	GOVERNMENT	105	CONCEPT	205	REVOLVER	305	PARENTS	405	NODULES
6	WORLD	106	SOCIETY	206	ROLE	306	ROCK	406	NORMAL
7	COUNTRY	107	NEED	207	ALL	307	STATEMENT	407	PRICE
8	STATE	108	SUPREME	208	ATMOSPHERE	308	TITLE	408	PROPER
9	ONE	109	AIR	209	FOOD	309	EARLIER	409	SCREEN
10	LAST	110	MEMBERS	210	PERIOD	310	FIFTH	410	TOPIC
11	SECOND	111	REST	211	ROAD	311	LETTER	411	TRAFFIC
12	INDIAN	112	AGE	212	SHEET	312	PARLIAMENT	412	YEARS
13	TIME	113	CITY	213	SOIL	313	ROOM	413	B
14	RIGHT	114	DECEASED	214	CAPITAL	314	BETTER	414	BALANCE
15	TWO	115	BOOK	215	CHAIRMAN	315	COMMON	415	BANKS
16	UH	116	FORMER	216	DATE	316	DATA	416	FIVE
17	WHOLE	117	GREAT	217	DOCTOR	317	DEFENCE	417	FRENCH
18	BEST	118	MOMENT	218	EIGHTH	318	QUALITY	418	GOVERNMENTS
19	THE	119	WESTERN	219	PAGE	319	TEMPLE	419	MALE
20	MOST	120	AMOUNT	220	SEA	320	VICE	420	NECESSARY
21	NEW	121	HOSPITAL	221	DECISION	321	ANSWER	421	OPENING
22	POLICE	122	AREA	222	ENVIRONMENT	322	DEN	422	OPOSITION
23	MAIN	123	DIFFERENCE	223	GAME	323	FOURTH	423	P
24	NEXT	124	MORE	224	LAW	324	FUNCTION	424	PLANET
25	WAY	125	PURPOSE	225	POST	325	INCOME	425	POSSIBILITY
26	END	126	ABOVE	226	SCENE	326	LIST	426	REASONS
27	STUDENTS	127	BLACK	227	TEACHERS	327	POSITION	427	SECRETARY
28	CASE	128	ENGLISH	228	WOMEN	328	REGION	428	SHORT
29	PRIME	129	LAND	229	YOUNG	329	STANDARD	429	STAGE
30	BALL	130	MEETING	230	MODERN	330	VALLEY	430	STRUCTURE
31	UNIVERSITY	131	VARIOUS	231	OBJECT	331	ACT	431	TEACHING
32	ONLY	132	BACKGROUND	232	AUTHOR	332	BRAIN	432	WEATHER
33	NATIONAL	133	COLLEGE	233	CHILDREN	333	CONTROL	433	WELL
34	PRESENT	134	CURRENT	234	DEATH	334	INITIAL	434	AUTHORITIES
35	ACCUSED	135	ISSUE	235	HISTORY	335	LATEST	435	BOTTOM
36	QUESTION	136	MAN	236	I	336	NATURAL	436	CHEMICAL
37	EARTH	137	SAID	237	INDIVIDUAL	337	ORDER	437	DEMAND
38	LEFT	138	OLD	238	NOVEL	338	RECENT	438	DIRECTOR
39	TOTAL	139	RURAL	239	PREVIOUS	339	RED	439	EXISTING
40	THIRD	140	IDEA	240	PROGRAMME	340	SECURITY	440	F
41	CENTRE	141	RATE	241	ACTUAL	341	STATUS	441	FRONT
42	FACT	142	FAMILY	242	AMERICAN	342	ACADEMIC	442	GOVERNOR
43	HOUSE	143	HUMAN	243	BOOKS	343	ART	443	GREEN
44	PROBLEM	144	LIFE	244	CHILD	344	BUS	444	LEADER
45	YEAR	145	ELECTION	245	CLASS	345	CONTEXT	445	MEDIA
46	CONGRESS	146	LINE	246	EXTENT	346	EXECUTIVE	446	MOON
47	MIDDLE	147	TERM	247	MOVEMENT	347	FAMOUS	447	PARTICULAR
48	NAME	148	A	248	NIGHT	348	IMPORTANCE	448	PICTURE
49	U	149	IN	249	PAPER	349	LONG	449	PLAN
50	PUBLIC	150	MAJOR	250	PART	350	MONTH	450	PROJECT
51	CENTRAL	151	POOR	251	PRIMARY	351	ORGANISATION	451	SCIENTIFIC
52	USE	152	BUDGET	252	REPORT	352	PRINCIPLE	452	SPEAKER
53	FOLLOWING	153	NATURE	253	STUDENT	353	SPECIAL	453	STATES
54	SITUATION	154	TOP	254	ARMY	354	TEMPERATURE	454	TRADITIONAL
55	THREE	155	WEST	255	BASE	355	TREE	455	COASTAL
56	WORK	156	PRESENCE	256	BORDER	356	WOMEN'S	456	COMMUNITY
57	BRITISH	157	RIVER	257	EFFECT	357	AFTERNOON	457	COMPANIES
58	CONSTITUTION	158	FUTURE	258	IMPORTANT	358	COMPUTER	458	CONFERENCE
59	PAST	159	HIGH	259	JANATA	359	EARLIEST	459	EAST
60	VERY	160	THING	260	MATCH	360	FINANCIAL	460	EDUCATION
61	WATER	161	LANGUAGE	261	METHOD	361	HIGHER	461	FULL
62	NUMBER	162	POWER	262	SOCIAL	362	INTEREST	462	INCREASE
63	DEVELOPMENT	163	DIFFERENT	263	TWENTY	363	LOWER	463	INJURED
64	ENTIRE	164	INTERNATIONAL	264	BANK	364	MAHARASHTRA	464	MESSAGE
65	SCHOOL	165	LEVEL	265	ECONOMY	365	QUESTIONS	465	PRIVATE
66	SUN	166	LIGHT	266	FINANCE	366	SIZE	466	PROGRESS
67	UNITED	167	NATION	267	PLANTS	367	SOLAR	467	RELATIONSHIP
68	WORD	168	RESULT	268	POPULATION	368	UHM	468	RULING
69	FINAL	169	SURFACE	269	REAL	369	WHAT	469	SELF
70	BASIC	170	STORY	270	CAUSE	370	ANT	470	SIMPLE
71	BODY	171	NORTH	271	CELL	371	BILL	471	SKY
72	COURT	172	RESULTS	272	GROWTH	372	BOY	472	SPIRIT
73	MATTER	173	SMALL	273	INFORMATION	373	DOOR	473	TECHNIQUE
74	DAY	174	TEACHER	274	LIBRARY	374	GIRL	474	TYPE
75	CHIEF	175	TEXT	275	PROBLEMS	375	GRASSHOPPER	475	APPLICATION
76	PARTY	176	BUSINESS	276	SO	376	HOME	476	AUDIENCE
77	COMPANY	177	PATIENT	277	UPPER	377	MIND	477	BIGGEST
78	GROUND	178	PRODUCTION	278	DISTANCE	378	NAMES	478	CORRECT
79	BLOOD	179	STUDY	279	HIGHEST	379	OCCASION	479	FOREST
80	DEPARTMENT	180	COURSE	280	NON	380	TIMES	480	FORMATION
81	FORM	181	ECONOMIC	281	PLANNING	381	WORST	481	GREATEST
82	HELP	182	EVIDENCE	282	SEMINAR	382	ABSENCE	482	IMPACT
83	PROCESS	183	HEAD	283	WORDS	383	BIG	483	INSTITUTE
84	BASIS	184	HONOURABLE	284	ANNUAL	384	CHANGE	484	LOK
85	HEART	185	PLANT	285	C	385	CLASSES	485	OFFICIAL
86	POINT	186	EVENING	286	HANDS	386	COLOUR	486	PAPERS
87	BEGINNING	187	MARKET	287	HOSTEL	387	FACE	487	PERSONS
88	FIELD	188	SIDE	288	JOB	388	LATTER	488	PROPOSED
89	FOUR	189	CYCLONE	289	LOCAL	389	MATERIAL	489	RAILWAYS
90	PLACE	190	INDUSTRIAL	290	MEANING	390	MAXIMUM	490	RESPONSIBILITY
91	PRESIDENT	191	KASHMIR	291	ORIGINAL	391	MOTHER	491	RISE
92	UNION	192	MONEY	292	OTHERS	392	RAILWAY	492	RULES
93	POLITICAL	193	BOARD	293	PROSECUTION	393	S	493	SEVENTH
94	GENERAL	194	GULF	294	REGISTRAR	394	TARGET	494	SHOP
95	NEWS	195	KIND	295	REMAINING	395	USUAL	495	SPOT
96	SOVIET	196	MANAGEMENT	296	SOUTH	396	ACTIVITIES	496	TERMS
97	COMMITTEE	197	PRESS	297	BOMBAY	397	BENEFIT	497	COLLECTOR
98	SENSE	198	PRINCIPAL	298	OO	398	BJP	498	COUNTRY'S
99	AWARD	199	THINGS	299	EARLY	399	DETAILS	499	EASTERN
100	SUBJECT	200	BACK	300	EARTHS	400	DISTRICT	500	EVENT

Top 500 R-1 collocates of *the* in ICE-KEN (unedited)

N	Word	N	Word	N	Word	N	Word	N	Word
1	SAME	101	BRITISH	201	INFORMATION	301	NON	401	FUND
2	GOVERNMENT	102	OPPOSITION	202	PARTICIPANTS	302	OFFICIAL	402	HIGHEST
3	THE	103	SPEAKER	203	POST	303	ORIGINAL	403	IMPACT
4	OTHER	104	NAME	204	WORDS	304	POEM	404	ISLAND
5	FIRST	105	PROCESS	205	CONTEXT	305	READER	405	MALE
6	COUNTRY	106	RESPONDENT	206	IMPORTANCE	306	SOUTH	406	MIDDLE
7	WORLD	107	APPELLANTS	207	MODERN	307	SUIT	407	SUN
8	MINISTER	108	YOUTH	208	POET	308	THING	408	BIGGEST
9	PEOPLE	109	CURRENT	209	POLITICAL	309	WIFE	409	BRAIN
10	WAY	110	CHIEF	210	POOR	310	DIFFERENCE	410	CAUSE
11	HOUSE	111	LATE	211	RECEIPT	311	DISCOURSE	411	COMING
12	TWO	112	GROUP	212	CAR	312	FATHER	412	COMPLAINANT
13	KENYA	113	INTERNATIONAL	213	DEPARTMENT	313	FRENCH	413	FOURTH
14	MOST	114	HEAD	214	DOCTOR	314	NORMAL	414	GIRAFFE
15	END	115	MARKET	215	ENTIRE	315	RIGHTS	415	GUY
16	FACT	116	MATTER	216	FARMERS	316	STATEMENT	416	INDIAN
17	TIME	117	POINT	217	FINAL	317	TOP	417	INSURANCE
18	WHOLE	118	STATE	218	RESPONDENTS	318	US	418	PARTICULAR
19	ACCUSED	119	AUDIENCE	219	COST	319	VILLAGE	419	PAYEE
20	PRESIDENT	120	COUNTRY'S	220	DOCUMENTS	320	AGE	420	PRODUCT
21	LAST	121	COURSE	221	FARMER	321	ATTORNEY	421	RECEIVING
22	MONEY	122	NAIROBI	222	KIPSIGIS	322	CHEQUE	422	RELEVANT
23	MINISTRY	123	PURPOSE	223	REGION	323	GREATEST	423	RULES
24	CASE	124	TOTAL	224	SECRETARY	324	INTEREST	424	SCHOOLS
25	SECOND	125	INDIVIDUAL	225	SENSE	325	KIKUYU	425	STANDARD
26	ASSISTANT	126	OF	226	SHOP	326	LAWS	426	VIDEO
27	POLICE	127	RURAL	227	BOY	327	LITTLE	427	WELFARE
28	PROBLEM	128	WORK	228	CONCEPT	328	MOTION	428	WOMEN'S
29	CHILD	129	COLONIAL	229	EARLY	329	POSITION	429	8
30	BEST	130	EFFECT	230	FLOOR	330	PREVIOUS	430	ABSURD
31	ONLY	131	MOMENT	231	GIRL	331	PRIMARY	431	AGRICULTURAL
32	UNITED	132	REASON	232	LONG	332	APPLICATION	432	ANIMAL
33	FOLLOWING	133	DEATH	233	RELATIONSHIP	333	ARTIST	433	CHIPS
34	NATIONAL	134	VERY	234	SCENE	334	BENEFIT	434	COUNTER
35	ONE	135	MESSAGE	235	WEST	335	BOYS	435	CROP
36	RIGHT	136	PRESENT	236	AMOUNT	336	BUDGET	436	DELIVERY
37	USE	137	MORNING	237	CENTRE	337	CHAIR	437	EFFECTS
38	AREA	138	OLD	238	CONVERSATION	338	CHEQUES	438	ENGLISH
39	SOCIETY	139	SOCIAL	239	HOME	339	CHURCH	439	EUROPEAN
40	LANGUAGE	140	NATION	240	JOB	340	CLASSROOM	440	EYES
41	MAIN	141	PRICE	241	ONES	341	COLD	441	INCIDENT
42	NEXT	142	VARIOUS	242	TEACHERS	342	DATE	442	INDUSTRIAL
43	FAMILY	143	WORD	243	BASIC	343	EXTENT	443	KITCHEN
44	QUESTION	144	MORE	244	DRIVER	344	HISTORY	444	MIND
45	CHILDREN	145	SOIL	245	LADY	345	INSTITUTE	445	MUSIC
46	NEW	146	IN	246	PROJECT	346	JUDGE	446	OFFICER
47	DAY	147	LEARNED	247	QUALITY	347	LATEST	447	PERFORMANCE
48	DISTRICT	148	PARTY	248	WRITTEN	348	MANY	448	PRESSENCE
49	APPELLANT	149	PLACE	249	CONFERENCE	349	MATERIAL	449	PRESSURE
50	HON	150	SECURITY	250	DECISION	350	PLOT	450	PRIVATE
51	UNIVERSITY	151	STORY	251	EVENING	351	PRACTICE	451	SONG
52	MAN	152	DISEASE	252	FOOD	352	PROSECUTION	452	STATUS
53	LOCAL	153	MEANING	253	PRESS	353	RULING	453	STRUGGLE
54	TEACHER	154	MEMBERS	254	SO	354	SHOW	454	SUCCESS
55	LAW	155	BLIND	255	TYPE	355	SKIN	455	TREASURY
56	CITY	156	CHAIRMAN	256	WORKERS	356	SYSTEM	456	VIEW
57	COMMUNITY	157	ECONOMY	257	BOARD	357	2	457	AIR
58	LAND	158	MEETING	258	CONSTITUTION	358	ACCOUNT	458	AREAS
59	PAST	159	BEGINNING	259	EDUCATION	359	ACTUAL	459	ASPECT
60	BABY	160	CENTRAL	260	LEARNER	360	BANK	460	BAN
61	MOTHER	161	LEARNERS	261	NORTH	361	COUNCIL	461	CHARGE
62	AFRICAN	162	MAJOR	262	PLAY	362	DIRECTOR	462	COMMISSIONER
63	PUBLIC	163	PROBLEMS	263	POSSIBILITY	363	HOUSEHOLD	463	CONDITIONS
64	NUMBER	164	STREETS	264	REASONS	364	HUSBAND	464	DISTRIBUTION
65	PROGRAMME	165	FUTURE	265	SPECIAL	365	KWS	465	DRUGS
66	UH	166	HOSPITAL	266	THEATRE	366	MACHINE	466	EAST
67	STUDENTS	167	PARENTS	267	ACT	367	NIGHT	467	ELECTIONS
68	KENYAN	168	ROOM	268	BOOKS	368	ORKOIK	468	FACE
69	REPORT	169	TRUTH	269	COMMITTEE	369	PASSBOOK	469	FINANCIAL
70	PERSON	170	2ND	270	CUSTOMER	370	STAFF	470	FIRE
71	DEVELOPMENT	171	BAR	271	DEGREE	371	UTERUS	471	IMMEDIATE
72	BOOK	172	BASIS	272	GOVERNMENTS	372	ADMINISTRATION	472	INDUSTRY
73	COURT	173	NEWS	273	HUMAN	373	AUTHOR	473	KNOWLEDGE
74	ABOVE	174	OTHERS	274	NEEDS	374	CLASS	474	LIST
75	ISSUE	175	STUDENT	275	REAL	375	CONTINENT	475	MECHANIC
76	ROAD	176	FOUR	276	RECENT	376	DEMAND	476	OCCASION
77	SCHOOL	177	THINGS	277	SAID	377	DIFFERENT	477	ORDER
78	TRADITIONAL	178	WATER	278	SMALL	378	ESTABLISHMENT	478	OVERALL
79	WOMAN	179	FORM	279	TOWN	379	FOREST	479	OWNER
80	EVIDENCE	180	FORMER	280	WESTERN	380	FULL	480	PAPER
81	IDEA	181	GROUND	281	WRITER	381	GREAT	481	PARTIES
82	WOMEN	182	MEN	282	WRONG	382	ISSUES	482	PROCEEDINGS
83	DECEASED	183	NATURE	283	BUSINESS	383	KANU	483	RESULT
84	DOOR	184	NECESSARY	284	CO	384	LESSON	484	STANDING
85	THREE	185	POPULATION	285	DEFENDANT	385	MILK	485	STREET
86	TRIAL	186	STORIES	286	KEY	386	NARRATOR	486	WHITE
87	REST	187	SUBJECT	287	PHYSICAL	387	OFFENCE	487	ANC
88	GENERAL	188	VICE	288	STUDY	388	PERMANENT	488	APPEAL
89	HIGH	189	WAR	289	TABLE	389	PRODUCTION	489	ARMED
90	ENVIRONMENT	190	ANSWER	290	TITLE	390	PROVISION	490	CIRCUMSTANCES
91	KIND	191	BACK	291	UN	391	REPUBLIC	491	DOCUMENT
92	BODY	192	COMMISSION	292	DEPUTY	392	RESPONSIBILITY	492	DRUG
93	MAJORITY	193	ECONOMIC	293	DISPLAY	393	TERM	493	DUPPLICATE
94	YEAR	194	FIELD	294	EXISTING	394	TURKANA	494	EXISTENCE
95	YOUNG	195	LEGAL	295	HEALTH	395	VEHICLE	495	HEART
96	NEED	196	PATIENT	296	INTRODUCTION	396	WINDOW	496	JUA
97	THIRD	197	PLAINTIFF	297	LATTER	397	ANIMALS	497	KBC
98	OFFICE	198	BUS	298	LEVEL	398	CHOICE	498	LACK
99	SITUATION	199	COMMON	299	MANAGEMENT	399	CIVIL	499	LEAST
100	ROLE	200	COMPANY	300	NATURAL	400	FEW	500	LIGHT

Top 500 R-1 collocates of *the* in ICE-PHIL (unedited)

N	Word	N	Word	N	Word	N	Word	N	Word
1	PHILIPPINES	101	STUDENT	201	SO	301	IMPORTANCE	401	MORO
2	FIRST	102	PROGRAM	202	TEACHER	302	SIX	402	NECESSARY
3	SAME	103	FUTURE	203	THINGS	303	SOCIAL	403	PACIFIC
4	OTHER	104	LANGUAGE	204	AMOUNT	304	APPLICATION	404	PERIOD
5	PRESIDENT	105	MOVIE	205	ATENEO	305	DATA	405	PRINCIPLE
6	THE	106	PATIENT	206	BOTTOM	306	FILIPINOS	406	RECENT
7	COUNTRY	107	DEVELOPMENT	207	COMMON	307	FOLK	407	TEMPERATURE
8	UH	108	STATES	208	COMMUNITY	308	HISTORY	408	DOCTOR
9	WORLD	109	GROUP	209	DEATH	309	INFORMATION	409	FRENCH
10	PHILIPPINE	110	IDEA	210	DEFENSE	310	KING	410	GLOBAL
11	GOVERNMENT	111	INTERNET	211	FINE	311	LATE	411	INITIAL
12	WAY	112	MONEY	212	GUY	312	ROLE	412	LETTER
13	ONE	113	SYSTEM	213	PROPOSED	313	SIZE	413	MEDIA
14	PEOPLE	114	LATEST	214	RULES	314	TREATY	414	MONTH
15	MOST	115	SITUATION	215	SMALL	315	EFFECTS	415	OPENING
16	UNITED	116	TOP	216	AGREEMENT	316	GENTLEMAN	416	OPPORTUNITY
17	TIME	117	TRUTH	217	BEGINNING	317	HONORABLE	417	PASIG
18	TWO	118	CURRENT	218	CHAIR	318	HOSPITAL	418	PRINCIPAL
19	U	119	PROJECT	219	COMMISSION	319	MEANING	419	QUEEN
20	SECOND	120	BIG	220	DECISION	320	OPPOSITION	420	RANGE
21	HOUSE	121	CHINESE	221	ENGLISH	321	OTHERS	421	RIVER
22	NATIONAL	122	JAPANESE	222	GAME	322	QUALITY	422	SERVICE
23	LAST	123	ONES	223	GIRL	323	TOPIC	423	WEATHER
24	SENATE	124	TOTAL	224	N	324	WEST	424	ABILITY
25	NEXT	125	AMERICANS	225	SPEAKER	325	ARMED	425	BUS
26	NEW	126	C	226	TEST	326	CONTEXT	426	CABINET
27	END	127	FIELD	227	WORK	327	FULL	427	CREATION
28	BEST	128	GENERAL	228	CENTER	328	LACK	428	DETAILS
29	RIGHT	129	AMERICAN	229	CHILDREN	329	RELEASE	429	IMPEACHMENT
30	ONLY	130	BASIS	230	FOOD	330	SOUL	430	NATURE
31	PAST	131	EARLY	231	KIND	331	SUN	431	STREETS
32	FILIPINO	132	RESULTS	232	PRODUCTION	332	YOUTH	432	SUBJECTS
33	WHOLE	133	BUSINESS	233	GREAT	333	BABY	433	TAX
34	DEPARTMENT	134	COMPANY	234	LEFT	334	DEAN	434	TRAINING
35	LDQUO	135	COURT	235	MAJORITY	335	EARTH	435	VISAYAS
36	FOLLOWING	136	FATHER	236	SAID	336	HIGHER	436	ABSENCE
37	RECORDING	137	FORMER	237	YOUNG	337	IMPLEMENTATION	437	ACTIVITY
38	CASE	138	MARKET	238	HEART	338	INTEREST	438	EDSA
39	PROBLEM	139	NEED	239	LIFE	339	LOWER	439	EL
40	FAMILY	140	SOUTH	240	MEANTIME	340	PLANE	440	GLOBE
41	FACT	141	EXECUTIVE	241	OBJECT	341	PRESS	441	HOLY
42	UNIVERSITY	142	FACULTY	242	PROPER	342	SECURITY	442	INDUSTRY
43	VERY	143	INTERNATIONAL	243	RED	343	WOMAN	443	LIST
44	NUMBER	144	LIGHT	244	SOURCE	344	A	444	NEEDS
45	PUBLIC	145	COURSE	245	DOLLAR	345	AGE	445	OWNER
46	NATION	146	LOWEST	246	FOURTH	346	B	446	PARTY
47	REST	147	PLACE	247	HIGH	347	CLASS	447	PASCHAL
48	STATE	148	CHILD	248	LONG	348	COMMELEC	448	PRIMARY
49	LAW	149	COMPUTER	249	MOMENT	349	EAST	449	PROVINCES
50	THREE	150	HUMAN	250	PREVIOUS	350	ESTRADA	450	R
51	USE	151	M	251	TITLE	351	POLICE	451	SET
52	V	152	VARIOUS	252	DOCUMENTS	352	RECORD	452	SOUND
53	THIRD	153	WOMEN	253	MAJOR	353	SIDE	453	SPECIAL
54	CITY	154	AVERAGE	254	MANILA	354	SPIRIT	454	STREET
55	QUESTION	155	ECONOMIC	255	OFFICIAL	355	STATEMENT	455	TEXT
56	REASON	156	GOOD	256	OIL	356	TABLE	456	USER
57	CONSTITUTION	157	I	257	POLITICAL	357	TOWN	457	YEARS
58	ENTIRE	158	REGION	258	REPUBLIC	358	WRONG	458	ABOVE
59	OLD	159	BASES	259	AFTERNOON	359	APPLICANT	459	ATMOSPHERE
60	P	160	MEN	260	BASIC	360	ASIA	460	BRAIN
61	MILITARY	161	POOR	261	LAND	361	BILL	461	CONSTITUTIONAL
62	POINT	162	THING	262	LINE	362	CATHOLIC	462	FINANCIAL
63	STUDENTS	163	AIR	263	PEACE	363	CAUSE	463	GREATEST
64	YEAR	164	FINAL	264	PRICE	364	CENTRAL	464	HOTEL
65	COMMITTEE	165	HEAD	265	ROOM	365	COMPLAINT	465	MASSES
66	PROCESS	166	MAN	266	SPANISH	366	EXTENT	466	NATURAL
67	OFFICE	167	MORNING	267	TRADITIONAL	367	GROWTH	467	OPEN
68	ECONOMY	168	PHONE	268	CONCEPT	368	MOTHER	468	OPPOSITE
69	MAIN	169	PRIVATE	269	FORM	369	NON	469	ORDER
70	MORE	170	SUPREME	270	GROUND	370	POSSIBILITY	470	PARENTS
71	WATER	171	USUAL	271	MANY	371	POST	471	POPULATION
72	NAME	172	ENVIRONMENT	272	ADMINISTRATION	372	TAPE	472	PRICES
73	POWER	173	NEWS	273	D	373	LHM	473	PRINCESS
74	STUDY	174	PESO	274	DOOR	374	UP	474	REGIONAL
75	DIFFERENT	175	MILF	275	EVENING	375	ART	475	STAGE
76	AREA	176	REAL	276	FRONT	376	AUTHOR	476	STOCK
77	SUBJECT	177	ASIAN	277	NET	377	DIRECTOR	477	TESTIMONY
78	ISSUE	178	BACK	278	PRESIDENTIAL	378	EVIDENCE	478	VALUE
79	LOCAL	179	IN	279	PROBLEMS	379	FILM	479	VOLUME
80	CHURCH	180	PROVINCE	280	PROSECUTION	380	MATERIAL	480	ACTUAL
81	PERSON	181	SECRETARY	281	REPORT	381	OF	481	ANTI
82	PRESENT	182	WORDS	282	S	382	ORGANIZATION	482	BANKS
83	WITNESS	183	COST	283	SENSE	383	PAPER	483	BETTER
84	HIGHEST	184	DIFFERENCE	284	TEACHERS	384	POLICY	484	BIDDING
85	DAY	185	FLOOR	285	WAR	385	RATE	485	CHALLENGE
86	WORD	186	MARCOSES	286	ANSWER	386	RELATIONSHIP	486	DARK
87	BUREAU	187	MIND	287	ARTICLES	387	STATUS	487	DATE
88	STORY	188	PURPOSE	288	CHAIRMAN	388	WIFE	488	DESIDER
89	US	189	FACE	289	COLD	389	CHANCE	489	DRUG
90	BODY	190	KEY	290	ELECTION	390	CLASSROOM	490	FREE
91	MEMBERS	191	MATTER	291	EXISTENCE	391	CROSS	491	NORMAL
92	SCHOOL	192	PART	292	FOREIGN	392	POSSIBLE	492	POTENTIAL
93	BOARD	193	WHAT	293	LEVEL	393	RAMOS	493	RIGHTS
94	MIDDLE	194	LATTER	294	PLAN	394	REGULAR	494	RULE
95	TERM	195	MARCOS	295	START	395	TYPE	495	SOLUTION
96	BOOK	196	NINETEEN	296	TWENTY	396	BACKGROUND	496	STRUCTURE
97	ORIGINAL	197	BALL	297	VISITING	397	CROWD	497	SUM
98	SEA	198	BIGGEST	298	BAR	398	INDIVIDUAL	498	SURVEY
99	BANK	199	PRESENCE	299	COLLEGE	399	L	499	VICE
100	FOUR	200	RESULT	300	EFFECT	400	LORD	500	ALLEGED