The Semantics of Optionality

by

Paul M. Heider

February 1, 2015

A dissertation submitted to the

Faculty of the Graduate School of the

State University of New York at Buffalo

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics

Acknowledgements

My pants got dirty. That's part of science. So is getting hurt and getting lost.

— Guthrie

This dissertation would be sorely lacking without the frequent and far-reaching help of my advisors Doug Roland and Jean-Pierre Koenig. Doug was willing to part with many of his precious hours discussing everything from crafting a good beer to crafting a good scientific story and all the statistics in between. JP gave me context and perspective to make sure I always knew which giants' shoulders I took a boost from, making my crafting that much more interesting and accurate. This story would be much more boring without them.

Cristin Kalinowski, Karl Sarvestani, and Kate Donelson also helped my story along by reading bad drafts, listening to half-formed ideas, and celebrating the minor victories after a long day of writing. Jon Bona and Marta Cieślak did wonders to keep me entertained at our regular disserdates. The ladies at Tom's Diner are in the unique position of having provided the vast majority of the caffeine that fueled my writing this past year. The Janya/ContentSavvy crew (especially Tom Cornell, Steve Duquette, Ray Wen, Laurie Crist, Jeremy Kornbluth, Harish Srinivasan, and Rohini Srihari) were very obliging to deal with my split brain while writing.

As with any story, context makes a difference and these people were crucial to my dissertation's context. Gail Mauner was a wonder to take me into her lab and share her world. Bill Rapaport (a gentleman scholar) and Stu Shapiro were also kind enough to further my cognitive science education in their lab. David Zubin was always happy to have a rambling debate or discussion (depending on the day). Colleagues and friends like Hongoak Yun, Erica Su, Sunfa Kim, Bret Bienvenue, Jeruen Dery, Cody Hashman, Stephani Foraker, Andreas Brocher, Eunkyung Yi, Jordana Heller, and Rui Chaves helped me develop a keener eye for experiments and analysis.

Then, there are the folks like Anne Colette Sheffer, Hsinwei Chen, Kirsta Mahonen, Fabian Rodriguez, Poornima Farrar, Yana Petrova, Albert Goldfain, Jason Naradowsky, Andy Wetta, Ashlee Wetta, Derry Moore, Emily Stewart, and Ben Madoff who helped me settle into my new academic home, complete with dinner parties and game nights. The Grinnell Plans community kept me connected to the world outside. Other folks who helped from afar and made me a better story teller in every visit include Jorin Garguilo, Alice Haltiwanger, Andrew Kaiser, Pete Brands, Chris Jones, and Marshall Derks. Mike Prentice doesn't fit easily into any one of the above categories because he really belongs in all of them.

Finally, Steve & Arvela Heider always made sure a certain starving grad student had his fill. Mom, Dad, Mary Winn, John, and all the four-legged bringers of peace and purring: you do and have done so much.

Contents

The author intended for this page to be blank.

List of Figures

The author intended for this page to be blank.

List of Tables

Abstract

For every participant role filler in an utterance, speakers must choose to leave it bare (e.g., "the interviewer") or to modify it (e.g., "the interviewer on Fresh Air"). Their decision is the end result of a combination of complex factors ranging from the original message to how distracted the speaker is. When we use corpora to create language models, part of our job is understanding the observable properties in and around an event description that allow us to predict these decisions. A considerable body of work on language production and discourse pragmatics concentrates on measuring noun phrase predictability and other forms of shared knowledge that help determine the balance point between over- and under-specification of a participant role filler. Although the importance of predictability as measured by long-term probabilities has long been recognized, I present a novel quantitative analysis of participant role filler predictability, the structure of the mental lexicon, and how the interaction of these two inform a speaker's internal perception of informativity. Standard Gricean assumptions tend to be efficiency oriented. Speakers will be informative enough but not wastefully so. Using these to model corpus distributions predict that noun phrase modification rates are directly proportional to predictability in order to satisfy the speaker's obligation to always be informative. In contrast, standard Firthian models (built around the idea that "you know a word by the company it keeps") assume spreading activation—and not efficiency—is the dominant predictor of usage. Sensitivity to activation's effect predicts that noun phrase modification rates are inversely propor-

tional to predictability. Strongly connected participant role fillers could be easily activated for production while weakly connected participant role fillers would either be mentioned less often or themselves trigger strongly connected features (not normally associated with the head verb) to be primed for production.

To distinguish between these competing assumptions, I analyze participant role filler modification rates in event descriptions with respect to three indicators: the syntactic and semantic optionality of the role filler, the general predictability of the verb's role fillers, and the predictability of individual pairs of verb/participant role fillers. First, I use insights from linguistic theory to classify verbs and their participant roles into classes of syntactic optionality and semantic optionality. Second, I quantify over a large corpus the general predictability of a verb's participant roles and the specific predictability of each pair of verb/participant role filler. Finally, I model the relationship between the three indicators and modification in order to ascertain whether speakers have a stronger tendency to modify the more predictable participant role fillers, as Grice's Maxim of Relevancy predicts, or a tendency to modify the less predictable participant role fillers, as a Firthian activation-based model predicts.

I present descriptive statistical models to chart the relationship between predictability, syntactic optionality of a participant role, and semantic optionality of a participant role. In general, verb classes with stronger mental lexicon connections to their participant role fillers according to theory also have more predictable participant role fillers in the British National Corpus. Specifically, syntactically optional direct object verbs and semantically obligatory instrument verbs have more predictable participant role fillers than the opposite, comparable verb class. I also present several linear mixed-effect models to determine how predictive of modification the independent variables of syntactic verb class, semantic verb class, and verb/participant role filler predictability are. According to these models, speakers are significantly more likely to modify the less predicted participant role fillers even when taking into account individual verb and verb class differences. I conclude that mental lexicon accessibility modulates noun phrase realization according to a Firthian activation-based model. For each factor, I discuss possible explanations for the correlations between modification, predictability, and optionality and how these correlations make sense within a larger production model.

The author intended for this page to be blank.

Chapter 1

Introduction

[B]revity is the soul of wit

— Polonius, *Hamlet*, Act II, Scene ii

The lady protests too much, methinks.

— Queen Gertrude, *Hamlet*, Act III, Scene ii

Imagine a dark and stormy murder mystery. A murderer uses a weapon to kill a victim. In describing this event, a speaker has several decisions to make. He can choose to mention the weapon or not, as in Example 1 taken from the British National Corpus[1] (BNC, Burnard, 1995) and adapted in Example 1a and Example 1b. He can also choose to omit the direct object (the victim), as in Example 1c. Whether or not he mentions these arguments, they are part of the event and are present in the speaker's internal representation of the event.

(1) In Ball, above, the accused killed *his neighbour* **with a shotgun**.

a. the accused killed *his neighbour* **with a shotgun**

[1] Examples cited below from the BNC were obtained under the terms of the BNC End User License. Copyright in the individual texts cited resides with the original IPR holders. The web site (http://www.natcorp.ox.ac.uk) contains further information and licensing conditions relating to the BNC.

b. the accused killed *his neighbour*

c. the accused killed **with a shotgun**[2]

Examples 1b and 1c provide less explicit information about the event. The speaker can also choose to be more informative about the event by providing additional information through modification about the victim or the weapon, as in Example 2.

(2) the accused killed ...

a. *his snooping neighbour ...*

b. **with a stolen shotgun**

Including this kind of information is always more informative to the listener than omitting it (compare the victims and weapons in Example 2 with those in Example 1). It is especially informative when the victim or weapon are of an unexpected type. We have to assume that speakers try to be informative, when possible, but that they also need to keep the message short enough to not lose the listener's attention or distract the listener from the core message. We, as users of corpora, serve as a special type of listener. We use what we can garner from a corpus to create a specific model of the intended message or a general model of the underlying language used. Creating a general language model from a corpus requires determining the baseline patterns of usage, explaining these baseline patterns, and then explaining deviations from these baselines. We need to understand the observable properties in and around an event description that predict baseline or deviant patterns. My dissertation investigates properties of event descriptions tied to predictability and informativity that explain people's choices under these circumstances.

Of course, not all event types are underlyingly equal. Some event types have generally much more predictable circumstances, participant roles, participant role fillers, etc. If we

[2] This particular example may sound a little odd in English. However, this potential awkwardness has more to do with the verb *kill* than with the construction itself. For instance, 'the accused dug with a shovel' is perfectly normal.

assume that the speaker is at least partially making her decision based on how informative she is being or could be, then different event types will have different baseline levels of informativity. Further, different real world events have different levels of predictability or novelty.

How do properties of the event type tied to the meaning of the verb influence how the event description is realized? For example, a killing event *sometimes* involves a weapon, whereas a shooting event *always* involves a weapon (and a specific type of weapon, at that). Interestingly, in neither instance (as in Example 1b, above, and Example 3, below) is the speaker grammatically obligated to mention the weapon.

(3) the accused shot his neighbour

In a similar vein, does the predictability of an argument's role filler change a speaker's decision to be more or less informative when describing an event? For instance, *coffee* is a very common direct object of the verb *drink* while *blood* is less common. A speaker may wish to emphasize *blood* over *coffee* because the latter is so mundane. However, novelty is not defined in terms of a fixed probability. Drinking *blood* is fairly expected for vampires and drinking *coffee* is unexpected when the speaker is talking about Mayans. Thus, there is a certain amount of situational context that impacts how surprising something is in an event description.

In using corpus data to categorize speaker behavior, we restrict ourselves to only that evidence which is available in the text itself. Pretend the subject and the verb are all a listener has heard so far about an upcoming event description. If he can fairly reliably guess any further role fillers for this description, then we can say that this event is predictable. If the likely role fillers come from too diverse a collection, he will have a hard time guessing correctly and we can say that this event is not predictable. Likewise, our corpus models can only make predictions based on the presence of particular words or phrases in the event description or the lack of a category of words or phrases. At the core of my previous questions about speakers decisions, we find the predictability of an event description given the verb. How this predictability correlates with the production decisions a speaker makes tell us about her preferred strategies for being informative, which helps us to determine reasonable baseline patterns of usage.

Predictability is an important concept because everything not explicitly said by the speaker must be guessed at or ignored by the listener. As a result, I treat informativity as the inverse of predictability. That is, the more something steers the listener away from his natural biases, the more informative it is.

A less informative variant can be used to describe a superset of the events described by the more informative variant. Example 1b, while being less informative, can still be used to describe the same event as Example 1a. It provides less information about the event, which means it does less to explicitly distinguish the intended event from other possible events. The event described in Example 4 is compatible with Example 1b but not Example 1a. As a result, the listener does not know for sure which event the speaker intends to describe when he hears Example 1b.

(4) In Ball, above, the accused killed his neighbour **with a knife**.

By not mentioning a weapon, the speaker forces the listener to come to his own conclusions about the presence or absence of a weapon in the event. If he assumes a weapon *was* used, the identity or general class of the weapon (e.g., a gun vs. a knife) must also be guessed. It may be quicker and easier for the speaker to omit the weapon, but omission could lead to confusion and a failed communication act. The communication act is a failure when the listener's understanding of the event deviates too much from the speaker's understanding.

The listener's understanding of an event can likewise be influenced by different levels of verbosity or detail about an explicitly mentioned event participant. More contextual support about an argument should provide a clearer picture than less contextual support. What if that additional information distracts the listener from the component of the event description that the speaker believes is more central or informative to the story? Because the crux of this study is comparing event descriptions in a corpus, I will focus my analysis on the speaker's decision to provide (or not) this contextual support for an argument. Should she modify an argument or leave it bare? Do we find modified or bare arguments? More importantly, are the argument modified when we expect them to be? What contextual evidence serves as the best predictors?

Modification, like omitted arguments, results in two variants with higher and lower levels of informativity, respectively. The speaker can modify the argument, and be more informative about the event she wishes to capture. Or, she can leave the argument bare, and be less informative about the details of the event. For instance, Example 5 describes another killing event from the BNC. I have bolded the phrase of interest.

(5) The court in Belfast ruled that Christie was more responsible for her actions when she killed Penny McAllister **with a sharpened butcher's knife** than was originally thought.

Providing additional information about the weapon is likely to increase the listener's agreement with the court's ruling because he knows more about the underlying event. Example 6 is a more extreme example of the same idea. The modification on the weapon provides details that bias us, as the listener, to assume that Christie had ill intent and did not accidentally kill Penny.[3]

(6) Christie killed Penny McAllister **with a dull, pen knife**

In contrast, the two unmodified variants in Example 7 present Christie in a more neutral

[3] This is the same conclusion that the court came to upon knowing all of the details of the event.

light (although perhaps not completely inculpable).

(7) a. Christie killed Penny McAllister **with a knife**

b. Christie killed Penny McAllister **with her car**[4]

Once again, modification requires more effort from the speaker, but results in a more informative event description. The speaker must choose between, on the one hand, the prolixity of mentioning and modifying every possible argument at the cost of extreme verbosity and, on the other hand, the parsimony of mentioning only the bare arguments at the risk of providing too little information to correctly convey his message.

When and how a speaker chooses to be informative tells us about her decision-making process. Understanding this baseline process helps us in turn to understand distributional patterns in corpora and build better language models. For instance, there is a speaker profile that I think of as exemplified by a traditional spreading activation model or a solicitous speaker. A speaker reporting an event with a predictable argument (e.g., the *car* in Example 8) will have no trouble mentioning the argument but is not particularly inspired to expand upon the argument. An unexpected argument (e.g., the *hammer* in Example 9) is sufficiently odd or unique on its own to draw attention to. The more of the speaker's attention an argument occupies while formulating the event description, the more likely she is to wax poetic about it in the final production. Alternatively, the speaker could feel the need to provide additional context about the argument, like a story-teller who wants to connect all the pieces in the puzzle for you. Predictable pieces are easy to integrate but unexpected pieces may require help. Either of these impulses would explain higher modification rates (i.e., increased informativity) in the context of less typical arguments than in the case of more typical arguments.

(8) All six admitted conspiring to steal cars.

[4] I assume that the possessive pronoun *her* does not qualify as modification. See Section 2.2 for the intuitions I use to define the labels "bare" and "modified".

(9) Hotek, a heretic priest of Vaul, steals the sacred hammer from Vaul's Anvil and makes his way to Nagarythe.

I caricature a second speaker strategy as a rational agent[5], an optimal agent, or, more colloquially, a numbers man focused on minimizing extra effort. A rational agent speaker reporting an event like that in Example 8 should more often either describe the event without the mention of the stolen object (because a listener would mostly guess it correctly) or expand upon the *car* to make that portion of the message contribute information about the event beyond the speaker's natural biases. Mentioning an unmodified direct object like *car* would be much less frequent than an unmodified direct object like *hammer*. Outside of the event described in Example 9, listeners do not expect to hear about a stealing event with a hammer involved. Mentioning that argument alone is probably enough to ensure a successful communication act. Any further information provided about the stolen object is informational icing on the cake.

These two speaker strategies can be used to discuss production choices and models of production. The strategies can also be used to ground our expectations of event descriptions in corpora and how we link those event descriptions with the underlying language models that generated them. For the remainder of this dissertation, I will contrast the Minimum Effort Hypothesis (i.e., the rational agent speaker) with the Maximum Context Hypothesis (i.e., the solicitous speaker). On the one end of the speaker strategy spectrum, the Minimum Effort Hypothesis predicts that more predictable contexts will more often be augmented or modified. On the other end of the spectrum, the Maximum Context Hypothesis predicts that less predictable contexts will more often be augmented or modified. Both approaches are viable production styles but one will tend to dominate over the other in naturalistic settings. To better understand the factors that influence our modeling expectation's balance point between these two extremes, I investigate several factors.

[5]This sense of "rational" is borrowed from game theory.

The first factor that I investigate is how predictable an argument is given the verb. Because the alternative to mention and modification (omission and leaving bare, respectively) forces the listener—and our models—to fall back to his natural biases, we might assume that the speaker would take into account the typicality of the argument when choosing whether to be more or less informative. For instance, the word *car* is one of the things most frequently mentioned as being stolen in the BNC, as in Example 8. In contrast, sentences about stealing hammers, as in Example 9 from the BNC, are far less common. How does the unexpectedness to the listener[6] of the stolen object affect the speaker's planned description? How does the unexpectedness change what else we expect to see (or not see) in the event description?

The next two factors, after argument predictability, that should be most influential to a speaker's production choice are tied to properties of the event type (i.e., rather than being tied to the particular event description). All arguments for an event type (e.g., *kill* vs. *shoot*) are classifiable as syntactically obligatory/optional and semantically obligatory/optional[7]. For instance, the direct object (an argument such as *his neighbour* in Example 1) of verbs like *kill* and *eat* is syntactically optional.[8] As shown in Example 1 and Example 10 (taken from the BNC), the direct object can be omitted and the event description is still syntactically well-formed.

(10) a. I ate *a fish with bones.*

b. I ate.

In contrast, verbs like *wear* and *hang* have syntactically obligatory direct object. The event

[6] I do not assume a strict mapping between a listener's real world expectations and their linguistic expectations. However, their real world expectations will be biased by (if not predominantly determined by) what they have previously been exposed to linguistically.

[7] See Section 1.1 for discussion of these categories.

[8] Unlike *eat*, *kill* is more marked when used without a direct object. The verb tends to reflect a habit (e.g., 'she kills' ≈ 'she is a killer' much like 'he steals' ≈ 'he is a member of the thieves guild') or requires a strong discourse context in the case of omission. See Section 1.1.3 for more discussion of these differences.

descriptions in Example 11 are acceptable while those in Example 12 are incomplete or ungrammatical.

(11) a. He wore *great clothes*.

b. The tailor hung *his creations* on the railings.

(12) a. * He wore.

b. * The tailor hung on the railings.

I should note that there is an interpretation of Example 12b that is grammatical and a complete event description. However, the event described by that sentence is very different from the event described by Example 11b. As such, the sentence in Example 12b does not represent a relevant alternation for Example 11b. A speaker who intends to describe the former event but uses the latter's description has failed her communication act. In contrast, either sentence in Example 10 is an acceptable description of the same underlying event and so, choosing between either alternate is a valid choice for the speaker.

Because the alternation is not available for all event types (due the idiosyncratic lexical properties of individual verbs), the range of possible grammatical descriptions is categorically different between syntactically optional direct object verbs and syntactically obligatory direct object verbs, as shown in Table 1.1. As a result, a speaker's choices are categorically different between the two verb classes. The different available choices, in turn, alter a listener's expectations.

Consider a speaker who uses a verb with a syntactically obligatory direct object to describe an event. He cannot save effort by omitting that argument, no matter how typical it might be. Under the Minimum Effort Hypothesis, mentioning an argument costs effort. Mentioning a patently obvious argument counts as a wasteful sunk cost. Mentioning a less obvious argument means that this additional cost is exchanged for less predictable information. Thus, his only option to make a syntactically obligatory and typical argument

Table 1.1: Range of Grammatical Descriptions by an Event Type's Syntactic Optionality

Direct Object Argument Type	Sample Event Description		
Syntactically Obligatory	(*unacceptable*)	He wore clothes	He wore great clothes
		(← *less informative* .	*more informative* →)
Syntactically Optional	I ate	I ate fish	I ate fish with bones
	(⟵ *less informative*	....	*more informative* ⟶)

not count as wasted effort is to make the argument more informative through modification. Sending good money after bad is usually considered a poor decision. However, speakers may be framing this strategy more in terms of adorning a necessary, but boring, structural component. In more concrete terms, we would expect highly predicted argument fillers to have higher modification rates for syntactically obligatory direct object verbs than for syntactically optional direct object verbs.

If speakers tend to be more solicitous in their event description choices, we should see a different pattern emerge. Under the Maximum Context Hypothesis, speakers will modify the less predictable argument fillers more often than the highly predictable fillers. Resnik (1993) has shown that syntactically optional direct object verbs have more predictable direct objects on average than syntactically obligatory direct object verbs. This general trend makes sense when you take the listener's perspective. The speaker can only reasonably omit a direct object if the listener can reliably recover the correct event description. It therefore stands to reason that verbs which license direct object omission have generally more predictable direct object arguments. For Maximum Context speakers, this trend means that, because syntactically obligatory verbs have less predictable direct object arguments, they will also have more often modified direct objects.

The semantic optionality of the argument is a second event type property that should strongly influence speakers' perception of predictability. The direct object in all the verbs used above is semantically obligatory. Even when the speaker omits *fish* in Example 10b's event description, it is still part of the underlying event. In other words, the fish was there whether the speaker mentions it or not. The participant denoted by 'the fish' was present in the event either explicitly (when mentioned) or by entailment (when omitted).

In contrast, the weapon of a *kill* event is semantically optional. Only some kill events have a weapon involved. The instrument of killing in Example 1 is a shotgun. If the shotgun had not been explicitly mentioned, it is still possible that it (or a similar additional entity) was part of the event. It is also possible that no such additional participant was part of the event. In this case, we want to think of the event as being without an instrument. If the accused had pushed her neighbor off of a cliff rather than shooting him, no instrument is semantically connected to the event.

Koenig et al. (2003) argue that semantically obligatory and semantically optional verbs have categorically different encodings for the same argument type. The common instruments for a semantically obligatory verb (or those instruments' properties) are partially encoded on the verb itself. To know what *cut* (a semantically obligatory verb) means involves understanding the action and the implement needed to allow the action to occur. In contrast, knowing the verb *kill* (a semantically optional verb) means understanding the action and perhaps having an idea of an implement that could facilitate the action. Hence, knowing the semantic optionality of an argument tells us a great deal about how strongly that argument (and its likely role fillers) are connected to the verb in question. Let us look to examples from the BNC to help shape good intuitions about semantic optionality.

In Example 13 from the BNC, a handkerchief fills the role of instrument. The handkerchief serves as this additional participant included in the event proceedings. Beyond this intuitive notion of instrument, I provide a technical definition of the term in Section 1.1.4. In Example 14 from the BNC, hands are used to do the rubbing. Following Koenig et al.

(2003), I assume that core body parts do not qualify as instruments if they are essential aspects of the event. The act of eating, in the common sense, is impossible without a throat. A throat cannot serve as a separate participant in the eating event as it is a core part of the eater. As such, this event description does not entail the use of an instrument. From these two examples, we can conclude that *rub* events sometimes entail an instrument and sometimes do not. It is a semantically optional instrument verb, like *kill*.

(13) Julia rubbed her eyes **with a handkerchief** and then went to retrieve her drink.

(14) Benjamin rubbed his face **with his hands**.

(15) Swod rubbed his knees with a groan.

(16) He rubbed his hands with satisfaction.

Further evidence for the semantic optionality of the instrument argument comes from Example 15. When unspecified, the listener assumes that Swod used his hand(s) to rub his knees. Even if his hand(s) is not directly rubbing the knee, it is presumably still gripping the actual rubbing implement. This bias to assume an unmentioned argument is actually a core body part is a good indicator that there is no semantic instrument required by the event. Swod could have alternately used a talisman or a wand to rub his knees. In these latter instances, the underlying event would entail an instrument. If Swod used just his hand(s), the underlying event would not entail an instrument. Additionally, this example shows us that the argument can be syntactically omitted. All the instrument verbs I investigate are syntactically optional instrument verbs.

The contrast to a semantically optional verb is a semantically obligatory verb. I will use the instrument verb *cut* to illustrate some of the important differences between these two verb types. Example 17 shows two events descriptions from the BNC. Both sentences contain an explicit mention of the semantic instrument used to perpetrate the cutting act.

(17) a. Auntie Eve cuts them **with her teeth**.

b. He cut off its head **with his sword**.

(18) a. Auntie Eve cuts them.

b. He cut off its head.

If the argument is omitted (as in Example 18, respectively), it is clear that an instrument is still required for the event to be complete. Example 19 and Example 20 show the absurdity of an instrumentless interpretation of the event described in Example 18. Returning to my original intuition of entailment of a new participant, we can see from these two examples that something above and beyond a normal pair of hands is required.[9]

(19) a. # Auntie Eve cuts them **with nothing**.[10]

b. # He cut off its head **with nothing**.

(20) a. # Auntie Eve cuts them **with her hand(s)**.

b. # He cut off its head **with his hand(s)**.

Our two extreme speaker types will, like with syntactic optionality, react differently to semantically optional and semantically obligatory instrument verbs. Semantically obligatory instrument verbs have generally more predictable instrument arguments. The most predictable instruments are (partially) encoded[11] on the verb, which makes them inherently less informative than semantically optional instruments.

I have outlined how the predictability of components in an underlying event can affect the production choices under the two hypotheses. Under the Minimum Effort Hypothesis, the speaker uses predictability to help her maximize informativity while minimizing

[9] To address an example brought up by Doug Roland, Jackie Chan's hands may well constitute such a novel entity. Those hands are especially apt instruments in the case of cutting things like heads. However, in this extreme case, we are also likely to talk about his hands as something requiring a license, just like a more canonical instrument—a weapon—requires.

[10] The verb *cut* can also mean *adulterate*. My analysis is focused on the core or primary sense of a verb. If Auntie Eve sells her drugs pure, then Example 19a **would** be a felicitous event description.

[11] I operationalize this partial encoding through using the synset at the convergence point across examples in my underlying representation of a verb's subcategorization. This partial encoding can also be operationalized in terms of necessary and/or sufficient features of the role fillers.

effort. Under the Maximum Context Hypothesis, the speaker uses predictability to help him gauge the parts of an event that likely need additional clarification. I use three predictors to determine which characterization of the production system seems to dominate in naturalistic speech: argument predictability, syntactic optionality of the argument, and semantic optionality of the argument.

I will show across all three predictors using naturalistic corpus data that speakers maximize informativity rather than simplify utterances. This preference supports the Maximum Context Hypothesis. Understanding this preference allows us to assume more accurate baselines for creating language models. These general results can be broken down into three major trends. First, highly predicted argument fillers have a lower modification rate (that is, the number of role fillers that are modified over the total number of role fillers) than their counterparts. Second, syntactically optional arguments have a lower modification rate than their counterparts. Third, semantically obligatory arguments have a lower modification rate than their counterparts. In general, the less predictable an argument filler is for an event, the more likely it will be modified by speakers. Critically, none of the individual three predictors can be reduced or simplified to the other two. If the predictors were reducible, my claims would necessarily need be likewise simplified.

I discuss these results in terms of two traditions of production and discourse research. The first follows from Firth's notion of "know[ing] a word by the company it keeps", which assumes a "mutual expectan[cy] and mutual prehen[sion]" between co-occurring words for the hearer/reader and speaker (Firth, 1968, pg. 179–180, 181). The Maximum Context Hypothesis is emblematic of this production paradigm.[12] Common, collocational pairs are strongly connected in the mental lexicon. One approach for resolving tip-of-the-tongue states and for aiding in memory retrieval is to activate words or memories

[12]Both of these traditions are relevant for understanding both production and comprehension paradigms. For the purposes of exposition in this dissertation, I focus on the production aspects of Firthian models. The same intutions should apply equally well to comprehension paradigms.

strongly connected to the elusive target. Spreading activation from the accessible words to the elusive target help with retrieval. In terms of event descriptions, the most difficult to retrieve targets are those arguments that are least predictable or least connected to the rest of the event description. As such, speakers may—intentionally or accidentally—be activating strongly connected concepts to the elusive argument in order to facilitate production. The nature of spreading activation means that these co-activated terms are also likely to be produced with the elusive argument. A second framing that connects a Firthian model to the Maximum Context Hypothesis is the more intentional strategy on the speaker's part of trying to provide bridging context. Namely, when a speaker is aware of weaker associations between the main event structure and a particular participant, she may choose to modify the participant in order to provide additional context to ease integration costs for the audience. This strategy requires more active intent from the speaker and thus would depend on her being able to afford the added processing load, which may not always be possible.

The other tradition follows from Grice's Maxims (Grice, 1975/2002, pg. 722), in which production is a calculated balance of brevity and informativity such that predictable parts of a potential utterance need to be augmented or omitted to maximize informativity while minimizing effort. The Minimum Effort Hypothesis is emblematic of this production paradigm. All else being equal, a highly predictable event description is less informative than a less predictable event description. One means for a speaker to increase the informativity of an event description would be to add more detail or modify an argument. Without that additional modification for otherwise predictable event descriptions, a listener may question what implicature he should make. Three of the four conditions that I investigate license argument omission. In these three cases, speakers may preferentially mention predictable arguments only when there is some additional, necessary modification to understanding the full story. In other words, we may also see higher modification rates on predictable arguments because the predictable arguments that did not warrant modification were left unsaid. The arguments alone were not sufficiently informative to satisfy the Maxim of Quantity.

The rest of this chapter covers the theoretical underpinnings of my study. In Section 1.1, I describe the argument types and verb classes that I will use to study speakers' modification choices. The full list of verbs and their classifications are in Appendix A. An index of verb mentions through the text can be found in the Index of Word Lemmas.

In Section 1.2, I approximate predictability and informativity used so far with models of entropy and relative entropy. Entropy is a measure of the uncertainty of event (i.e., the speaker's choice of event description) in terms of all possible outcomes (i.e., all valid, grammatical, and true event descriptions). Relative entropy is a measure how much a preceding event (e.g., choosing a verb) changes the uncertainty of a second event (e.g., choosing an argument filler).

I use noun phrase realization as an indicator of informativity and accessibility (i.e., strong connection in the mental lexicon or salient presence in the current discourse model). In Section 1.3, I discuss previous work that also draws connections between these three concepts. Much of the previous literature focuses on whether the noun phrase is a pronoun or headed by a full noun. My interest in noun phrase realization is primarily in terms of noun phrase type. I assume informativity and accessibility are important determiners of whether a head noun is modified or bare.

1.1 Participant Roles and Role Fillers

Until now, I have intuitively used the terms argument, direct object, and instrument. More precisely, I am concerned with the participant roles associated with a verb (and the event type it represents) and the role fillers that realize those roles in a particular event description. In Chapter 3 (my first series of studies), I analyze the various participant roles that are realized syntactically as the direct object of the verb. Direct objects provide an interesting domain of study because they can be syntactically obligatory or optional, depending on the verb. This optionality distinction should provide insight into two categorically different types of connection in the mental lexicon. In Chapter 4 (my second series of studies), I analyze the instrument participant role. Instruments likewise provide an interesting domain of study because they can be semantically obligatory or optional, depending on the verb. This slightly different optionality distinction provides contrasting insight to that provided by the direct object types of connections.

Again, in both cases, the *possible* role fillers for these roles vary by verb. This variation becomes a categorical distinction that we can leverage to learn about predictability and speaker's choices. For instance, only certain things can be killed. That which cannot be killed is not a possible role filler for the patient of a *kill* event. The *particular* role filler depends on the event description. For instance, the *kill* verb in Example 1 has a realized direct object. The role filler in this event description is a *neighbour*.

Some of these participant roles are core (obligatory) properties of the event, which leads to certain implications for organization of the mental lexicon. The patient of a *kill* event, for instance, is core to the meaning of the event. No body, no murder. The agent of a *kill* event (i.e., the murderer) is also core to the meaning of the event. For my purposes in creating comparison classes, I assume that this semantic coreness of the roles implies that there is a static connection in the mental lexicon between the event type and the roles. I use the term static because the connections cannot be removed without inherently changing the event. Additionally, activating an event type in order to describe an event provides easy access to any important features of the roles, including likely role fillers.[13]

[13]Those features may be directly encoded in the lexicon/lexical definition of the role or they may be otherwise encoded in some aspect of world knowledge. Either possibility equally supports my further claims regarding accessibility between concepts. What matters for my investigation is that certain event

In contrast, semantically optional participant roles for a verb (e.g., the instrument of *kill*) are not necessarily required to have these same static connections back to the event type, which also leads to certain implications for organization of mental lexicon. The connection between role and event type could be activated by some automaticized process in the speaker's production system, some intentional strategy of the speaker, or only when explicitly present in the speaker's understanding of the underlying event. In any of those cases, the participant roles (and their role fillers) are not as automatically or essentially accessible from a verb's mental lexicon entry as when the role is core to the meaning of the verb.

When a particular event is being described, the core roles and any pertinent peripheral roles will be associated with role fillers. For instance, the abstract role of murderer and victim are at least each associated with an individual. Depending on the world knowledge of the speaker, those individuals may be completely unspecified (e.g., the killer is still unknown) or fully specified (e.g., the speaker did it). Those connections between the abstract roles and concrete role fillers are dynamic connections, important at the time of production but not part of the speaker's long-term mental lexicon. I defer discussion of these dynamic connections and how to determine them until Section 1.2.

In the following sections, I define the static connection types associated with verbs and their participant roles. These definitions in turn will help us understand how participant role types impact accessibility and play a part in the greater question of role filler realization.

types are associated more strongly with certain participant roles, which are in turn differentially associated with certain participant role fillers.

1.1.1 Distinguishing Core from Peripheral

Core and peripheral properties of a verb have different types of connections in the mental lexicon. As I mentioned in the last section, core participant roles should have stronger, more reliable connections to their verbs than peripheral participant roles. For my purposes, the distinction between core and peripheral differentiates those things which *must* be encoded in a lexical entry and those things which may be, but are not necessarily, encoded. As I am interested in the relative informativity of a speaker's options, the necessity of encoding or lack thereof serves as another source of input data to the speaker.

Since at least Tesnière (1959), researchers have tried to distinguish between more core and more peripheral participant roles. The clear intuition repeated by scholars across time and frameworks (e.g., Bresnan, 1982; Foley & Van Valin, 1984; Pollard & Sag, 1987; Dowty, 2003) is that the event description is dependent on the core roles and can optionally be augmented by the inclusion of peripheral roles. I follow the more specific definitions of Koenig et al. (2003), which follow.

A verb's core roles are obligatory, while its peripheral roles are optional. I implicitly treat these labels as categorical distinctions. However, some theories either do not directly contrast them (e.g., MacDonald et al., 1994) or treat them as gradient properties (e.g., Manning, 2003; Rissman, 2010, 2013).

The participant role fillers *student* and *coffee* in Example 21 occupy the unarguably core roles of stirrer and stir-ee. But what of other roles such as instrument (e.g., Examples 21a and 21b), event location (e.g., Examples 21c and 21e), event time (e.g., Examples 21d and 21e), and participant location (Example 21e)?

(21) The grad student stirred her coffee

a. …with a spoon.

b. …with a pen.

c. …in the lounge.

d. …after the lab meeting.

e. …in the lounge after the lab meeting in her favorite mug.

Some of *stir*'s roles are syntactically obligatory (e.g., the stirrer and stir-ee), while some are syntactically optional (e.g., the instrument and the location). All of these roles are semantically obligatory but only some of them are necessarily lexically encoded (e.g., the instrument vs. the location). Location specifics could be encoded on the lexical entry for this event type but most of those restrictions would equally well apply to all events in a place with oxygen and gravity, etc. The instrument restrictions, however, are much more specific to this event type (or, at the least, a small class of events).

This finer level of role specification constitutes part of how a verb's meaning is encoded in the mental lexicon. Core syntactic roles are called complements, while peripheral syntactic roles are called modifiers. In a similar vein, core semantic roles are described as arguments, while peripheral semantic roles are described as adjuncts. Knowing what the verb *stir* means in English requires knowing which participant roles are complements vs. modifiers and which roles are arguments vs. adjuncts.[14] Table 1.2 depicts this contrast using *bring*, *eat*, and *stir* as examples.

I should be clear that lexical specification (i.e., roles licensed by a verb's mental lexicon entry) is not the same as lexical mention (i.e., roles included in a given utterance). Syntactically optional participant roles, like the direct object of *eat*, must still be specified in the verb's mental lexicon entry because it is semantically obligatory even if it does not always get mentioned in an utterance. Selectional constraints on syntactically optional roles are a strong a-theoretical justification for why obligatory roles must be specified in the mental lexicon. Selectional constraints restrict the possible fillers for a participant role. If

[14]Again, some of this semantic information may not be exclusive stored in the mental lexicon but rather part of general world knowledge.

Table 1.2: Comparable Participant Roles for Each Optionality Category

		Syntactic	
		Obligatory	Optional
Semantic	Obligatory	direct object of *bring*	direct object of *eat*; instrument of *stir*
	Optional		instrument of *eat*

a syntactically optional role filler is subject to selectional constraints, those constraints must be specified somewhere. On the one extreme, a strongly lexicalized theory allows these restrictions to be directly encoded on the the verb's lexical entry. The verb's entry would include the necessary semantic information required to define the selectional constraints. On the other extreme, a lexical entry minimally includes any explicit information about selectional constraints. Most selectional constraint information would be derived from incidental or strategic relations or connections that the verb happens to participate in. Since the presence or absence of a connection from a verb to any other construct is the primary determiner of selectional constraint, these connections would need to be rich enough to allow arbitrary types and arbitrary granularities of constraint. For example, the verb *devein* would somehow need to be connected in a way that restricts its direct objects to *shrimp*.

The presence or absence of direct objects is quite easily assessed. Levin (1993), for instance, attempts to exhaustively list all the different syntactic alternations of every English verb. I will be using Resnik (1993)'s judgment for my syntactic verb class distinction and leave further discussion of the direct object participant role to Section 1.1.3. Distinguishing arguments from adjuncts, however, is less of a settled issue. I cover my basic tools for this task in the next section. My final list for the semantic verb class distinction comes from Koenig et al. (2003) and Section 1.1.4 provides a more in-depth discussion of the instrument participant role.

In order to clarify the underlying differences between semantically obligatory and semantically optional participant roles, I will discuss the criteria for distinguishing arguments from adjuncts in this section (for an overview of the criteria, see Koenig et al., 2003). The main intuition agreed upon across theories is that argument information is crucial to the lexical entry of a verb, while adjunct information is extraneous to the lexical entry. The verb (as partially determined by its underlying event type) require a certain number of participants, which fit certain criteria (e.g., animacy, count, etc.). Other participants can be mentioned but they are not verb-specific or restricted by the verb. Koenig et al. (2003) formalize this intuition in the Lexical Encoding Hypothesis as follows:

The Lexical Encoding Hypothesis (Koenig et al., 2003, pg. 75, original emphasis)

> A participant role is a (semantic) argument of a verb if and only if it satisfies both the [Semantic Obligatoriness Criterion (see below)] and [Semantic Specificity Criterion (see below)], that is, if its presence is required of all situations described by that verb *and* if it is required of the denotation of only a restricted set of verbs.

This hypothesis, in turn, is dependent on two more formalized criteria: the Semantic Obligatoriness Criterion and the Semantic Specificity Criterion. First, according to the Semantic Obligatoriness Criterion (repeated below), a participant role must be part of all events described by the verb. That is, any event without that participant role is infelicitous.

The Semantic Obligatoriness Criterion (Koenig et al., 2003, pg. 72)

> If r is an argument participant role of predicate P, then any situation that P felicitously describes includes the referent of the filler of r.

Second, according to the Semantic Specificity Criterion (repeated below), the participant role is only licensed by a fraction of all possible verbs and events types. That is, not just any event-type licenses the participant role.

The Semantic Specificity Criterion (Koenig et al., 2003, pg. 73)

> If r is an argument participant role of predicate P denoted by verb V, then r is specific to V and a restricted class of verbs/events.

In assuming argument and adjunct status is encoded in the lexical entry, I am assuming the status is inherently verb specific. I can compare speakers' modification decisions within participant roles, between verbs of different semantic or syntactic optionality, and between participant role types. The following section describes the particular participant roles I plan to investigate and the verb classes that license them.

1.1.2 Participant Role Types of Interest

I have chosen to focus on the participant roles that are encoded by the direct object and the instrument participant role. I will be able to use these two variables to juxtapose syntactic optionality (via direct objects, as per the columns in Table 1.2) with semantic optionality (via instruments, as per the rows in Table 1.2). Both distinctions should have an impact on predictability and modification rates. These distinctions are a matter of framework or infrastructure. Without a strong theoretical bent, it is underdetermined as to whether strong connections due to syntactic optionality will behave contrastively or in parallel with strong connections due to semantic obligatoriness. Contrastive effects on production choices imply that semantic and syntactic optionality differ qualitatively in their behavior. Parallel effects on production choices imply that semantic and syntactic optionality differ only quantitatively in their behavior. In the next sections, I will describe the general properties associated with obligatory and optional participant roles for each participant role type. Knowing these general properties will help justify predictions with respect to the role types.

1.1.3 Verbs with Direct Object Participant Roles

Some verbs syntactically require a direct object while others optionally allow for its omission. However, different verbs vary with respect to the degree of omission that they allow. For example, Lehrer (1970) divides direct object omissible verbs into four categories based on the preconditions of that omission: Type I–IV. For Lehrer's Type I verbs, the direct object is highly predictable because the verb that licenses it has strong semantic requirements defined in the mental lexicon. The high predictability warrants that they need not be mentioned (e.g., *read* clearly requires something readable and *eat* requires something edible). An omitted direct object is either directly recoverable or its most important properties are recoverable. Levin (1993) roughly classifies the same set of verbs as Indefinite Object Alternation verbs. These verbs have direct objects which Fillmore (1986) would say can be omitted indefinitely. They neither need be mentioned in the discourse nor be known to the speaker, as in Example 22.

Fillmore's contrasting class of definite omission need not be mentioned but must be known to the speaker, as in Example 23, usually because they have been sufficiently recently introduced in the discourse so as to be available.

(22) He ate (although I'm not sure what he ate).

(23) He called (#although I'm not sure who he called).

Lehrer classifies these as Type III verbs. Resnik (1993), following Cote (1992), calls these verbs Specified Object Alternation verbs. In Example 24, *Kent* can be omitted from the response because it is clear from the discourse that he would be the recipient of the phone call.

(24) a. Can you tell Basil to call Kent?

b. He called (Kent) already.

Lehrer's Type I and Type III verbs are intended to be mutually exclusive. The same phenomenon of omission is tied to two different sets of conditions: a single incredibly predictable filler or previous discourse mentions that makes a single filler incredibly predictable.

Type II verbs are the results of a mixture of the conditions for Types I and III. Type II verbs have two or more common direct object role fillers, like *answer* in Example 25 (Lehrer, 1970, pg. 240). These fillers are sufficiently different to not "guarantee unique recoverability" (Lehrer, 1970, pg. 242) but still be restrictive enough to allow for very strong guesses. For instance, the discourse context alone (without taking into account previous mentions) can sometimes be enough to guarantee recoverability.

(25) a. answer_1: letter, cable, card, ...

b. answer_2: question, request

A final relevant class includes Lehrer's Type IV verbs (e.g., *steal* or *interrupt*). They permit object deletion despite having looser selectional constraints on the direct object. To contrast Lehrer's Type I and Type IV verbs, Type I verbs include much of the information content critical to their direct objects, at least in the form of selectional constraints. Type IV verbs, however, do not include as much information content because the selectional constraints are fairly loose. Thus, these verbs do not encode as much information about the event because the basic properties of the direct object are too loosely defined.

Levin (1993) claims that the implicit assumption of omission is that the most frequent direct object is activated when no direct object is present[15]. She contrasts syntactically optional verbs like *eat* with syntactically obligatory verbs like *brush*. When the direct object of *eat* is not mentioned, a frequent filler (or, at least, salient features of frequent fillers) can easily be conjured to fill that particular empty role. However, *brush*-style verbs do not have a single clear direct object to fill that role. There is no clear super-ordinate category or hypernym for the direct object of *brush*-type verbs that does not also include unacceptable direct objects. However, both she and Resnik (1993) admit a certain incompleteness to this analysis. That is, predictability of role filler seems to be a necessary condition of syntactic optionality, but it is not a sufficient condition. Verbs like *devein* and *diagonalize* (as in Example 26) accept only a single class of filler each, but both syntactically require the direct object (McCawley, 1968). In Chapter 5, I sketch how verb telicity could be a necessary contributing factor to explaining why verbs like these, in addition to verbs like *devour*, are *not* syntactically optional.

(26) a. He deveined the shrimp.

b. She diagonalized a matrix.

For my purposes, it is relevant to note that this theme of direct object filler predictability arises several times when discussing syntactically optional direct object verbs (see, for instance, Hankamer, 1973; Sag, 1976; Williams, 1977). Resnik (1993) confirmed in a smaller corpus study that this hypothesis based on intuition matches the empirical data for thirty four of the most common direct object verbs in the Brown Corpus (Francis & Kučera, 1982). Theoretically, these results imply that syntactically optional direct object verbs encode at the least essential features of their most frequent direct object fillers. Will the same result hold for a larger corpus (with an inherently more diverse genre of event de-

[15] Again, this primarily relates to Lehrer's Type I verbs, although some Type II verbs fit this description.

scriptions) and for a larger set of verbs (because we have already discussed ways in which many of these properties are idiosyncratic to a verb)? I interpret my results to imply that syntactically optional direct object verbs have strong mental lexicon connections to the most predicted fillers. What we do not know is whether predictability of filler also correlates with higher or lower modification rates of the filler itself. Speakers have clearly been influenced by the predictability of the fillers diachronically to allow them to become syntactically optional roles. Is there evidence that speakers will also be influenced by predictability on a shorter-term scale (i.e., within a discourse) to alter their modification decisions as a result? Chapter 3 addresses these questions in more detail. As a preview of the answers, syntactic optionality continues to be strongly related to selectional constraint (η) in a larger corpus and with a larger set of verbs. Extending the verb list to include lower frequency verbs seemed to weaken the relationship but did not remove it altogether. I also find that predictable role fillers correlate with lower modification rates, further implying that speakers are indeed impacted by this predictability even within a discourse.

1.1.4 Verbs with an Instrument Participant Role

Some verbs *semantically* require the instrument participant role while others optionally allow for instruments. Instruments make for an interesting case study in predictability exactly because of this split in semantic optionality. We, as listeners, only sometimes assume an instrument is part of an event. If we have assumed an instrument and one is explicitly provided in an event description, then we learn a certain amount of information intended to be conveyed by the speaker. If we have not assumed an instrument and one is explicitly provided, then we learn even more information. This baseline informativity difference should tell us a great deal about the speaker's intents and decision making process. I will focus on those parts of previous research into instruments and their use that most directly inform our understanding of event description predictability.

First, I will discuss the semantic criteria used to determine if a participant role filler is, in fact, an instrument. From these examples, we can generalize as to whether the verb semantically allows or requires an instrument. Finally, I discuss how this analysis of what constitutes an instrument plays into my working concept of a mental lexicon.

Many sources (e.g., Nilsen, 1973; Brunson, 1992; Fillmore, 1968; Jackendoff, 1987) provide details about what does and does not constitute an instrument. I will be assuming the primary intuition of the "used X to"/"Ved with X" alternation (cf. Lakoff, 1968; Koenig et al., 2008). Namely, is one participant being used by another participant to perform an action? If one participant could be realized as either the prepositional phrase in one frame (as in Example 27) or the direct object of *use* in a "used X to" frame (as in Example 28), it should be considered a verb that can take an instrument role.

(27) The `SUBJECT` `VERB`ed the `DIRECT OBJECT` with `X`.

(28) The `SUBJECT` used `X` to `VERB` the `DIRECT OBJECT`.

A second heuristic, to be met in addition to the first heuristic for verifying instrumenthood, is that the subject using the instrument should have agency over the instrument.[16] For instance, the chimney in Example 29 does not have any control over the smoke that it is filling the house with. In contrast, we would normally assume that the child in Example 30 does have agency over her actions.

(29) The chimney filled the house with smoke.

(30) The child filled the bowl with cream.

These criteria were most important for developing my pilot data, as described later in Section 2.4. As described in Chapter 2, I do not employ sufficiently fine-tuned filtering

[16] This restriction is semantically similar to Lakoff's (1968, 8 ex. 12) "purposive action" requirement.

on my final corpus results to distinguish in my qualitative results between the types of *with*-phrases contrasted by Example 29 and Example 30. Because of the potentially different semantic roles contrasted in Example 29 and Example 30, I wanted to first pilot my analysis using more uniform data (i.e., all examples included agency). Manual inspection of a subset of the data leads me to conclude that there were not enough subjects without agency over the instrument to negatively impact my overall findings. Even if there had been a significant number of subjects without agency, the possible differences in underlying event types or differences in semantic role connections (namely, which roles were truly optional or obligatory) would decrease the chances of finding clear results.

Based on the alternation criterion and agency criterion, we can then determine whether the instrument role for the verb in question is semantically obligatory or semantically optional. In other words, is it essential to the understanding of the event that an instrument be used? For instance, in Example 31, both *poke* and *sip* can take instrument roles (as shown in Example 32).

(31) a. The police officer poked the body (with a stick).

b. The police officer sipped his iced tea (with a straw).

(32) a. The police officer used a stick to poke the body.

b. The police officer used a straw to sip her iced tea.

However, the instrument is lexically required (read: entailed) in Example 31a but is not lexically required in Example 31b. We will call *poke*-style verbs *obligatory instrument verbs* and *sip*-style verbs *optional instrument verbs*.

Koenig et al. (2008) provides a thorough survey of all English verbs with respect to the semantic optionality of their instrument role. All verbs which allow an instrument participant role, whether obligatorily or optionally, will be called instrument verbs in later discussions in this dissertation. Appendix A and Lexicon 4.1 (in Chapter 4) list all the instrument verbs from Koenig et al. with a BNC frequency of at least 10. Verbs with fewer sentence samples were dropped from my analysis in order to protect against extrapolating verb trends from too small a sample. There is no theoretical reason why these omitted verbs should be disqualified from my qualitative conclusions about the different classes.

As I mentioned in the last paragraph, not all instrument verbs are analyzed in this dissertation due to the practical issue of insufficient data. I also do not analyze those instrument participant role fillers that surface through a different alternation than the prepositional "with" structure that I describe in Section 2.3. For instance, instruments can also occur in the subject position (e.g., Example 33) and in the direct object position (e.g., Example 34).

(33) Rosewater makes a good aftershave.

(34) The apothecary used rosewater to make a good aftershave.

Finally, as another conservative analysis practice, I only considered non-phrasal verbs as instrument verbs. Phrasal verbs (as in Example 35) that otherwise met the criteria to be analyzed as instrument verbs were ignored because the syntactic annotation of these examples make them particularly prone to miscategorization.

(35) He cut out her heart with a hunting knife.

Instrument verbs occur in corpora with a realized instrument at a much lower rate than direct object verbs and a realized direct object. Interestingly, the mention rate of instrument role fillers does not appear to be significantly different when comparing semantically obligatory and semantically optional verbs (Koenig et al., 2003, pg. 91; Yun, 2012, pg. 65). It would not be surprising if semantically obligatory *instrument* verbs and their instrument fillers have equally strong connections in the mental lexicon as syntactically optional *direct object* verbs and their direct object fillers. Both verb classes have *syntactically* optional and *semantically* obligatory fillers. Semantically optional instrument verbs are the most likely to pattern differently from the other three. This final class of verbs have a *syntactically* and *semantically* optional filler of interest, meaning connections between these verbs and their instrument role fillers should be the weakest.

The syntactically optional *direct object* verbs and both classes of *instrument* also cluster together because all three groups have syntactically optional fillers. This syntactic optionality means that the Minimum Effort Hypothesis is not as extremely tested as with the syntactically obligatory *direct object* verbs. Namely, speakers using a verb from this fourth class are syntactically obliged to include the role filler. In case of the first three verb classes, the speaker may not be syntactically obliged to include the role filler. Hence, an argument based on recovering sunk costs through modification is not quite as strong. Nonetheless, the highly predicted role fillers should leave the speaker with a context-neutral sense of roteness. Even when other linguistic or metalinguistic factors bias the speaker to produce the role filler, role fillers more likely to be produced given the verb will always feel less novel and less interesting than an unlikely role filler.

1.2 Modeling Accessibility and Informativity

We need a measure to approximate relative connection strength between verbs and role fillers. Knowing how strongly connected verbs and role fillers are will tell us how informative a particular role filler is for a given event type. Strict co-occurrence frequency is not a good indicator of relative connection strength. A particular role filler could have incredibly high counts with a particular verb but only because it is a generally frequent word rather than because it is relatively significantly associated with the verb in question. For instance, pronouns occur very frequently with all verbs but that does not mean that they are more strongly associated with any of these verbs. Instead, following Resnik (1993), we want to focus on the *entropy* of participant role fillers and *relative entropy* for those fillers being chosen with respect to a particular verb in an utterance.

Relative entropy can be broken down into two events: the given event and the event that follows it. How much does knowing the given event change the outcome of the following event? If the two events are fairly unrelated (e.g., the chance of drawing a card given that you rolled a die), then the relative entropy is low. The distribution of cards that could be drawn is unrelated to how the die landed. Each card has the same probability of appearing on top before and after the roll. If the two events are somehow dependent on each other (e.g., the chance of drawing a card given that you already drew ten cards from the same deck), then the relative entropy is high. The distribution of cards that could be drawn will change dramatically based on which cards were already drawn. The probability of some cards appearing on top will go to zero because they were already drawn from the deck. The probability of other cards will slightly increase because there are ten fewer alternate cards. In this instance, drawing the eleventh card from the deck introduces slightly less information than drawing the first card because, at the eleventh draw, there are fewer possible outcomes.

In the context of my studies, our given event is the mention of a particular role filler. Our following event is the mention of that same role filler given a particular verb. How often does a role filler occur associated with the verb in question as a function of how often it occurs overall?

The word *coffee* occurs as a direct object of the verb *drink* 92 times in the BNC. That means 14.3% of *drink*'s direct objects are *coffee*. If we think knowing the verb *drink* has relatively little to do in changing our expectations about relative frequencies of direct objects, then we have to assume that *coffee*, in general, occurs as the direct object of about 14.3% of all verbs in the BNC. Clearly, 1 out of every 7 direct objects across all verbs are not *coffee*. Thus, *drink* and *coffee* have high relative entropy. Specific pairs of verb/participant role filler have a relative entropy value that is also called selectional strength. A verb's overall relative entropy (i.e., the sum of its selectional strength with each of its role fillers) is called both η and selectional constraint.

In his dissertation, Resnik (1993) showed a strong correlation between the strength of a predicate's selectional constraint and how often that predicate omits its direct object in several corpora. Resnik's bridging assumption is that predicates with very strong selectional constraints must have those constraints encoded in their lexical entry. Because these entries have a strong bias for only a subclass of direct object role fillers, the direct object could be omitted if it fell within the category of most likely filler. The direct object *pizza* is omissible in Example 36 because the possible object role fillers for *eat* are dominated by food items.[17] More generically, the direct objects of *eat* can be clustered into a coherent class.

(36) Hank ate [pizza]

In contrast, *pizza* is not omissible in Example 37 because a wide range of word classes occupy the object role of *bring*. A coherent class for the direct object's of *bring* does not seem to exist. An alternate analysis is that *bring* does not allow direct object omission because we (as speakers, as a culture, etc.) never care about *bring*-ing events without also caring about what was brought. However, comparing the example sentences in Example 38 provide evidence that something more is at stake. In both instances, the social responsibility of contributing to the group meal is important. With the verb *donate*, the listener can extrapolate that Hank contributed something of value (e.g., money or food). With the verb *bring* and the very specific context of a potluck, the listener should be able to extrapolate that Hank brought food but something blocks omission of the direct object.

(37) Hank brought pizza

[17] Like the verbs *devein* and *diagonalize*, *devour* seems an ideal candidate for syntactic omission. I sketch a hypothesis in Chapter 5 why *devour* is *not* syntactically optional.

(38) a. * Hank brought for the potluck.

b. Hank donated for the potluck.

The direct objects *hair* and *teeth* are likewise not omissible in Example 39 because of the high frequency of both concepts without a clear super-ordinate term that encompasses both without including unacceptable direct object participant role fillers (e.g., *body part*).

(39) Harvey brushed...

a. his teeth

b. his hair

Below, I give a brief overview of Resnik's measure of selectional constraint (defined for a predicate) and the related measure of selectional association (defined between a predicate and a role filler). I propose to use these definitions of selectional constraint and selectional association as measures of the general predictability of a verb's role fillers and the specific predictability of a particular role filler given a verb. In the following sections, I will clarify several definitions relating to the notions of informativity and predictability, as well as lay down some groundwork about my working model of the mental lexicon. I will also explain the mathematical concept of entropy and relative entropy, which I use to quantify selectional constraint and selectional association.

1.2.1 Entropy

Entropy measures the uncertainty of an event. For this dissertation, the events we care about are the mentions of particular verbs, direct object concepts[18], and instrument concepts. As per Shannon (1948), uncertain events are, on average, very informative because we have few preconceived biases for the outcome. Whatever the actual outcome

[18]We can easily map from lexical items to concepts by substituting any occurrence of a word with the WordNet synset (Fellbaum, 1998) it is associated with.

is, we need to adjust our world model because we could not have reliably predicted the results. Other events (that is, events with highly predictable outcomes) are, in contrast, *un*informative. The outcomes of these events fit within our preconceived biases. For example, the constant drone of an airplane engine is uninformative and highly predictable exactly because it continuously communicates to us that the engine is working properly and is our expected outcome of riding in an airplane. In contrast, the sudden explosion of the same engine would be informative and unpredicted because it is highly unexpected. Linguistically, a parallel can be made to the occurrence of *the* and *sesquipedalian* in a sentence, respectively.

Relative entropy measures the uncertainty of one event with respect to a second event. If the occurrence of the given event does not change the expectedness of the outcomes of the subsequent event, then the former event is uninformative as to the likelihood of the latter event. Previous high predictions for the outcomes will remain high and previous low predictions will remain low. Linguistically, a parallel can be made as to the informativity of seeing *devein* or *shrimp* with respect to each other. The verb *devein* is a syntactically obligatory direct object verb and, to the best of my and McCawley (1968)'s knowledge, *shrimp* are the only things that can be *devein*ed. Seeing *devein* informs us with a certainty of 100% that we will see *shrimp* (or a pronominal reference to *shrimp*) in that same sentence, as per Example 40.

(40) Caroline **deveined** $\left\{ \textit{shrimp} \right\}$.

However, seeing *shrimp* does not equally guarantee that we will see *devein* in the same sentence, as per Example 41.

(41) Caroline $\left\{ \begin{array}{r} \textit{ate} \\ \textit{cleaned} \\ \textit{deveined} \\ \textit{had} \\ \ldots \end{array} \right\}$ **shrimp**.

Because the verb *devein* is fully predictive of the concept *shrimp*, seeing *shrimp* after seeing *devein* is uninformative. It provides no new information. The probability that *shrimp* will follow *devein* is one. In contrast, seeing *shrimp* before seeing *devein* is informative because there were other possibilities. The probability that the verb is *devein* given *shrimp* is less than one (because of *ate*, *cleaned*, etc.) and greater than zero.

To get beyond basic intuitions about entropy, we need to look at the mathematical formulation of the measure. Entropy is defined as the function $H()$ over n events with probabilities of $p_1, p_2, \ldots, p_n$, as per Equation 1.1.

$$H = -K \sum_{i=1}^{n} p_i * log(p_i) \tag{1.1}$$

The maximum certainty arises when a single event is guaranteed to occur. The probability of that event (say, p_1) is one, while all other events have a probability of zero. As shown in Equation 1.4, the entropy for this occurrence would be zero, matching our description of the event of choosing a direct object to follow *devein*.

$$p_1 = 1 \tag{1.2}$$

$$p_2, \ldots, p_n = 0 \tag{1.3}$$

$$H = -K(1 * log(1) + 0 * log(0) + \ldots + 0 * log(0)) = 0 \tag{1.4}$$

In contrast, the maximum *un*certainty arises when every event is equally likely. As shown in Equation 1.6, the entropy for this occurrence is normalized to one by means of the constant K.

$$p_1, p_2, \dots, p_n = \frac{1}{n} \tag{1.5}$$
$$H = -K\left(\frac{1}{n} * log\left(\frac{1}{n}\right) + \frac{1}{n} * log\left(\frac{1}{n}\right) + \ldots + \frac{1}{n} * log\left(\frac{1}{n}\right)\right) = -K * log\left(\frac{1}{n}\right) \approx 1 \tag{1.6}$$

No obvious analog to this occurrence exists in natural language, but one can imagine events with more equal distribution among their possible outcomes than the previous example.

The more important mathematical concept for this dissertation is relative entropy. Namely, how much does the identity of one event change the expected outcomes of a subsequent event? Instead of entropy's $H()$ function, the relative entropy of p with respect to q is defined as the function $D()$ over events p and q, as per Equation 1.7.

$$D(p||q) = \sum_{x} p(x) log \frac{p(x)}{q(x)} \tag{1.7}$$

Notably, the $-K$ constant is absent. As we are only concerned with relative values of entropy, all constants can be factored out. Otherwise, the equation looks strikingly similar to that of entropy in Equation 1.1.

Within the context of this dissertation, our two events of interest are the speaker's choice of participant role filler (q) and his choice of that same role filler given the verb in question (p). Resnik (1993, pg. 58), describing the same pair of events, summarizes relative entropy as a measure of "the cost of not taking the predicate into account." High relative entropy correlates with high selectional constraint for a verb. That is, the expected outcomes for the participant role filler knowing the verb are significantly different from the expected outcomes for that same participant role filler without knowing the verb.

Using relative entropy to measure the strength of predictability between a verb and each of its participant role fillers allows us to easily formalize two measures: selectional strength and selectional constraint (η). Selectional strength is the pairwise strength of association between a particular verb and a particular participant role filler. The stronger the selectional strength, the more the verb biases the occurrence of the the participant role filler. The value η (a formalization of selectional constraint) is then the sum of selectional restrictions of a particular predicate across all participant role fillers. Looking at Equation 1.8, we can see how η is explicitly formulated.

$$\eta_v = \sum_c p(c|v) log \frac{p(c|v)}{p(c)} \tag{1.8}$$

In this equation, η_v is the selectional constraint for the verb v. The independent event is $p(c)$, or the probably that the role filler (or concept) c is used in the given participant role. The depending or following event is $p(c|v)$, or the probability that the role filler c is used with the verb v. Note in this equation that we are summing over all possible role fillers (all c's) for the verb v.

Predicates with high η values (high selectional constraint) are fairly picky about their participant role fillers, while predicates with low η values (low selectional constraint) occur more often with a larger variety of role fillers. I have included a toy example of entropy and relative entropy in Appendix B. I also relate this toy corpus to a sample of real examples from the larger corpus.

Now that I have covered the formal definitions of informativity and predictability, I will discuss their implications on noun phrase realization. I will return to these formulas in Section 1.2.1 when I detail how to automatically determine these values from a corpus.

1.3 Accessibility, Informativity, and Realization

Semantic accessibility and perceived informativity lie at the heart of this dissertation. In the spirit of Collins & Quillian (1969, inter alia), I will assume the mental lexicon is a multi-dimensional network of hierarchical concepts, linguistic categories, and word forms. Other network-based models of the mental lexicon should be equally acceptable replacements to a Collins & Quillian model. An explicitly network-based model simplifies the mappings from the model to my definitions of probability and accessibility. Nonetheless, I have avoided describing my questions and analyses in a wholly framework dependent manner. Decentralized or distributed models (i.e., those lacking a single, core lexicon such as Elman (2009)'s Dinosaur Bones Model) can be adapted to integrate my results.

The primary sense of accessibility I use throughout can be grounded in Collins & Quillian's network-based metaphor. To co-opt a phrase from Hebbian learning (Hebb, 1949), *concepts* that fire together, wire together. Two words that co-occur in an utterance share a link. That link can be strengthened by regular re-occurrence until it becomes a permanent component within a person's mental lexicon. That link can also decay from lack of use. In other words, a highly accessible word (or concept) has strong connections in the mental lexicon to currently activated words (and concepts).

Perceived informativity (i.e., how informative a listener thinks something is) can be cast as the inverse relation to accessibility. A speaker will assume something to be informative if it contradicts her own assumptions about likely continuations. She knows her intended message and, consequently, the components of that message that most differ from baseline assumptions (i.e., generally unlikely lexical and syntactic choices) and context-specific assumptions (e.g., unlikely lexical and syntactic choice given other components of the utterance). Mentioning high probability concepts is less informative than mentioning low probability concepts, as a less frequent concept more strongly counters her interlocutor's assumptions. I am intentionally not assuming that my production model measures the informativity of each individual lexical choice. The metric I use abstracts away from lemmas to WordNet synsets to simulate a speaker being sensitive to the novelty of their target concepts. I assume that the general predictability or novelty of a concept is considered by a speaker when deciding what parts of the greater message to include, to focus on, etc.

In Section 1.3.1, I discuss how accessibility in the discourse and in the common ground have been shown to impact realization in the lab and in more natural settings. In Section 1.3.2, I talk about how informativity changes a speaker's production choices[19] and her interlocutor's comprehension performance. The examples in these two sections lay out important theoretical and mechanistic groundwork for framing the core problem in this dissertation. Namely, how is noun phrase realization impacted by strongly predicted and weakly predicted sentential contexts? How does the accessibility and informativity of a speaker's intended message affect his final choice in realizing that message? Specifically, I will look at participant role fillers, the noun phrases that denote the participant roles of a predicate.

1.3.1 Accessibility and Realization

The primary type of accessibility that I am concerned with is accessibility within the mental lexicon. Previous work looking at over- and under-specification of noun phrases has focused on discourse accessibility, shared common ground, or physical co-presence with the interlocutors. Greater accessibility in all of these domains has been shown to affect realization.

[19] Production is, of course, much more than simple accessibility and informativity. An utterance is more than just the most accessible words and concepts in the speaker's mind at any given time. It must realize the speaker's intended message.

As one exemplar of previous work in this area, Chafe (1972, pg. 50, footnote 3) defines *foregrounded* as "an abbreviated label for *assumed to be in the hearer's consciousness*" (emphasis in the original). Foregrounded items are treated differently from discourse new items by being phonetically realized with "low pitch and amplitude" or "weak pronunciation" (pg. 51). The lexical unit can also be (optionally) pronominalized. Foregrounding is assumed to be a necessary (but not sufficient) condition for pronominalization, in general. As the speaker and hearer are always in the "hearer's consciousness", they are always foregrounded. This permanent foregrounded status allows them to be felicitously referred to pronominally at any time.

Gundel et al. (1993)'s Givenness Hierarchy provides a similar, more formal framework for the mapping between accessibility and realization. The cognitive status of a nominal referent is conventionally signaled by its surface form. Following Table 1.3, the givenness is along the top row with English examples[20] along the bottom row. If a referent meets all the requirements to be realized as more than one of the forms in the table, the left-most form will be preferred.

Table 1.3: The Givenness Hierarchy (Gundel et al., 1993, pg. 275)

in focus	>	activated	>	familiar	>	uniquely identifiable	>	referential	>	type identifiable
{*it*}		{ *that* *this* *this* N }		{*that* N }		{*the* N }		{ indefinite *this* N }		{*a* N }

Likewise, a core tenant of Centering Theory (Grosz et al., 1983, 1995) is that pronominal realization is correlated with the focus of attention. The *attentional state* modeled by this theory is "an abstraction of the focus of attention of the participants as the discourse unfolds" (Grosz & Sidner, 1986, pg. 175). From a production perspective, a referent

[20]This hierarchy was tested on English, Japanese, Mandarin Chinese, Russian, and Spanish.

currently contained in the attentional state is more likely to surface as a pronoun. From a comprehension perspective, referents currently contained in the attentional state are candidates for co-reference with pronouns in an utterance.

More recently, Arnold (1998)'s dissertation analyzes how recency and subjecthood of a referent (along with three factors less relevant to the current discussion) impacts its realization and whether that referent will surface again later in the discourse. The more recent or focused a referent is, the more likely it is to be in pronominal form.

Other researchers have focused on why a speaker chooses to be verbose rather than why she chooses to reduce her utterance. van der Sluis & Krahmer (2007), for instance, used a point-and-name task in Dutch to look at referential form. The less accessible targets, because they were physically more distant, resulted in realizations that over-specified the referent. That is, the referent was realized in a more verbose than required form. This kind of very iconic accessibility altered speakers' productions.

1.3.2 Informativity and Realization

Informativity, much like accessibility, has been shown to impact speakers' production choices. We can think about one of the primary goals of modification as disambiguating a referent. If a speaker chooses to disambiguate a term by adding modifiers to it, it suggests that there is something in the contrast set for it to be confused with. In other contexts, a speaker may use modification to provide descriptive or subjective content (e.g., "a black dog" or "a dangerous shark", respectively, Dale & Reiter, 1995, pg. 234). The listener, in either case, knows that any modifiers introduced by the speaker are additional to the syntactically minimal message.

More concretely, speakers sometimes contrastively describe targets even when the targets are unique in the current context because of lingering contrast with previous, similar targets (Levelt, 1989; Spivey & Richardson, 2008). This effect of neither discourse present nor physically present competitor objects on referential choices may be similar to the distance-based effects found by van der Sluis & Krahmer (2007).

Brown (1985, and see also Brown & Dell, 1987) found that speakers, when re-telling a story, are more likely to mention atypical instruments than typical instruments. Much like Brown, I look at participant role fillers (i.e., the noun phrases that denote the participant role). However, instead of analyzing the binary mention/absent split, I use the bare/modified dichotomy. Specifically, I look at how role fillers are realized in strongly predicted and weakly predicted contexts to gain insight into how the two notions of accessibility and informativity impact a speaker's production decisions.

Most surprisingly from Brown's results, this mention disparity exists even when the audience to the story was already made aware of the atypical instrument. A strong interpretation of informativity's effect on mention would assume that once the atypical instrument was made known to both parties, it would lose its special status as highly informative. As a result, mention rates should go down. Instead, Brown found that speakers are unable to fully control this reflex. Their actions do not distinguish between the general case of the speaker knowing the identity of the atypical instrument prior to production and the specific case of both the speaker and audience knowing the identity of the atypical instrument prior to production. In other words, it appears that a speaker's reaction to atypicality is grounded more in the general atypicality of a role filler rather than the specific informativity in the current context of mentioning a role filler.

In a picture naming task, Engelhardt & Ferreira (2012) found that speakers acoustically produce the same noun phrase differently depending on whether the string contained a single source or multiple sources of disambiguating information about the objects present in the context. The observed acoustic reduction was discussed in the same light as syntactic reductions related to accessibility.

Interestingly, we find similar effects in comprehension. Namely, interlocutors are expecting this balance between parsimony and prolixity. When size (and other percept) modifiers are used (Sedivy et al., 1999), listeners assume that this additional information is being used with the intent of disambiguation. This assumption leads to anticipatory eye-movements for listeners to objects that have that perceptual property even before the head noun is known or identifiable (see also, Altmann & Kamide, 1999; Sedivy, 2003, 2006; Arts et al., 2011; Horowitz & Frank, 2012; Grodner & Sedivy, 2011).

As I mentioned before, all of these studies focus on external measures of predictability or accessibility that are active in the discourse or in the physical world. What about measures internal to a speaker? The most relevant study above is the typical/atypical instrument retelling by Brown (1985). Regardless of the actual discourse context (i.e., whether or not the audience knew the instrument's identity), speakers were biased to mention atypical instruments because their atypicality, in the larger picture, is normally informative.

We can specifically frame the major problem in this section in terms of maximizing informativity or maximizing accessibility. Namely, are speakers biased to maximize informativity through modifying highly predicted participant role fillers more often than highly unpredicted role fillers? Maximizing informativity satisfies the general intuition that speakers should try to be informative whenever possible and not provide obvious information without reason, as per the Minimum Effort Hypothesis.

In terms of Grice's Maxims (Grice, 1975/2002, pg. 722), good interlocutors are brief when possible, explicit when necessary, but always informative. In other words, a speaker will plan her utterance to be informative according to her internal expectations (i.e., those in her mental lexicon) and interpersonal expectations (i.e., those from the real world and discourse context).

Or, are speakers biased to maximize accessibility through providing more context for low probability role fillers than for high probability role fillers? Maximizing accessibility could help ease their production or could help their interlocutor more easily understand the underlying message, as per the Maximum Context Hypothesis.

Firth (1968, pg. 179–180) is often indirectly quoted for his lexicographer's stance that "[y]ou shall know a word by the company it keeps".[21] The common interpretation of his claim is that words will typically co-occur with other words most closely related to them. Put another way, words always tend to occur with another strongly related or associated word. Unmodified and weakly predicted role fillers are left without this strongly related word. Modifying them allows the speaker to provide that extra context or association.

For the rest of my dissertation, I focus on the notion of informativity between a verb and its direct object or instrument participant role fillers as determined by relative entropy and its derived measures. How much relative entropy is there between a particular verb and a particular participant role filler (i.e., selectional strength)? How much relative entropy is there for a verb between it and all its participant role fillers (i.e., the η value)? Do we find the same correlation between a verb's selectional constraint (η) and its syntactic optionality class as Resnik did when we use a larger corpus with more verb types? Or were his results an effect of using the most frequent syntactically optional verbs?

Finally, we can reframe the conflict between the Minimum Effort Hypothesis and the Maximum Context Hypothesis in terms of selectional strength. Does high selectional strength/constraint bias speakers to be more informative and modify the otherwise highly predictable participant role fillers in an utterance? Or does low selectional strength/constraint bias speakers to boost accessibility and modify the otherwise surprising participant role fillers in an utterance?

Chapter 2 focuses on the technical aspects of answering these question through col-

[21]This quote is commonly adopted by computational linguistics under the name of the Distributional Hypothesis.

lecting verb/argument pairs from the BNC and using these pairs to generate predictability measures (following Resnik, 1993). I formalize my syntactic definitions of bare and modified role fillers that serve as the critical dependent variable in my statistical models. I also describe general model properties, consistent across all of my statistical models used in later chapters, at the end of the chapter.

Because syntactic optionality and semantic optionality are conceptually different, I have separated my investigation into two major studies. Chapter 3 analyzes the impact of argument predictability and syntactic optionality or obligatoriness of the direct object argument on speakers' direct object modification choices. Chapter 4 presents a similar analysis for instrument predictability and the semantic optionality or obligatoriness of the instrument argument.

I will show in Chapters 3 and 4 that speakers tend to boost accessibility when describing events. That is, less predicted role fillers are significantly more likely to be modified than highly predicted roles fillers. These trends are true whether we compare individual verbs using the selectional strength for verb/role filler pairs or across verb classes split by the semantic or syntactic optionality its argument. These results support a stronger Firthian influence (the Maximum Context Hypothesis) than a Gricean influence (the Minimum Effort Hypothesis) on production.[22] Interestingly, these trends also show that selectional strength is more important than a verb's selectional constraint (η) for predicting modification.

Chapter 5 discusses the implications of my results for theory-based and data-driven models of semantic relatedness. These results are generally of interest to anyone looking at information processing and the effects of speaker design/audience design on production choices. Work in this area tends to focus on the external cues to production choices

[22] I do not refute a Gricean influence on production models. The data merely supports the claim that the Firthian component is stronger, when in conflict with a Gricean component.

while my results provide important baselines related to internal cues (namely, how informative role filler choices are outside of the current discourse context). Additionally, my results are of interest to anyone using corpora to model semantic similarity or relatedness. The core of all of these models is some variant of the Firth's Distributional Hypothesis. My results improve our understanding of what should be considered a baseline for co-occurrence between verb, role fillers, and those role fillers' modifiers. Finally, I discuss further avenues of research sparked by my investigations.

The author intended for this page to be blank.

Chapter 2

Analysis Methods

Words, words, words

— Hamlet, *Hamlet*, Act II, Scene ii

You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant.

— Sherlock Holmes, *The Sign of the Four*, pg. 196

This chapter focuses on the technical aspects of my investigation into speakers' choices. I detailed in the last chapter how we can use noun modification as a proxy for whether a speaker wishes to be more informative than his baseline minimal requirement (e.g., the range of event descriptions in Table 1.1). In order to compare his production choices, I needed a parsed corpus containing naturalistic event descriptions across a wide range of contexts. When he modifies a noun phrase, I assume he is trying to be additionally informative. My studies, in particular, focus on the context in which this potential modification occurs. Does the primary verb in the given event description tend to have fairly predictable role fillers or are they fairly unpredictable? Does the verb/role filler pair have a high selectional strength binding them together in the mental lexicon or are they more weakly associated? I also needed a sufficiently formalized definition of informativity to allow us to automatically measure high and low values per event description. Two of these measures are determined by corpus distributions. I include more details on how to convert corpus distributions into these metrics of selectional strength and selectional constraint. The third metric is a categorical property of the verb with respect to the role filler of interest.

In order to elucidate potential differences introduced by the type of role filler, I split the larger analysis of the effects of informativity and predictability on noun phrase realization into one set of studies focused on direct object role fillers and another set focused on instrument role fillers. In the direct object role filler studies, I compare modification rates between syntactically obligatory and syntactically optional direct object role fillers. I confirm the hypothesis that the distinction between syntactically obligatory and optional role fillers correlates with weaker and stronger connections between the verb and role filler in the mental lexicon, respectively. Because a speaker's internal understanding of informativity should be partially determined by her perceived strength of association in her own mental lexicon, this categorical distinction could be an important predictor of production choices. In contrast, instrument role fillers are always syntactically optional. Instrument verbs do differ in whether the instrument is semantically obligatory or semantically optional. Like with syntactic optionality, semantic optionality also correlates with strength of connection in the mental lexicon. As such, I compare modification rates between semantically obligatory and semantically optional role fillers in the instrument studies. Because of the overlap in methodologies, I analyze the effects of verb/role filler pairwise predictability on noun phrase modification in both studies.

The dependent measure in both the direct object and instrument studies is the binary classification of role fillers as bare or modified. In Section 2.2, I describe the semantic intuitions behind the labels "bare" and "modified" and provide the basic syntactic definitions for the labels “bare” and “modified”.

Evidence for these studies is taken from event descriptions. The primary predictors of interest are tied to the (un)expectedness of a role filler, which requires us to develop a measure of expectedness. By looking at the regularity of co-occurrence between verbs and their arguments, we can build a picture of what the average listener is exposed to.[1] Event descriptions from naturalistic speech give us evidence for the final production choices that speakers made in trying to convey their message. Syntactic analysis of the descriptions can tell us when a speaker had other options or if his hand was forced. In Example 11b (repeated below), the direct object is obligatorily included but unmodified. Hence, the speaker had two options: to modify the role filler or leave it bare.[2] He does not have the option to omit the role filler. In Example 2 (repeated below), the weapon is optionally included and modified. Hence, the speaker had three options: 1) to omit the weapon, 2) to include it and leave it bare, or 3) to include it and modify it.

(11b′) The tailor hung *his creations* on the railings.

(2′) the accused killed his snooping neighbour **with a stolen shotgun**

2.1 Corpus Details

As I am interested in comparing production choices in high and low predictability environments, I need to analyze event descriptions with realistic production patterns and a range of topics. I will use a balanced corpus with naturalistic sentence productions from the British National Corpus (BNC, Burnard, 1995). The BNC is composed of 100 million

[1] These expectations reflect their linguistic expectations for event descriptions but not necessarily their world knowledge expectations.

[2] I am simplifying the decision process a bit by describing it in this manner. I describe the decision as if it occurs at the final, downstream endpoint at which is matters rather than the upstream, higher-level point at which it actually occurs. The speaker is not deciding the noun to use and *then* deciding whether to modify it. In reality, his decision to modify or not is something closer to the emergent effect of deciding how much and what type of detail to include about a participant role filler’s underlying concept.

words split between written (90%) and spoken (10%) documents. The written genres include newspapers, periodicals, journals, books (both fiction and non-fiction), letters, and essays, among others. The spoken sources are informal dialogue ranging in register from radio shows to government meetings.

In order to allow for syntactic tree-based searches, this version of the BNC has further been automatically parsed by the Charniak Parser (Charniak, 1995, 1997). These tree-patterns allow me to isolate the corpus sentences containing the particular verbs and noun phrase role fillers of interest. I can also automatically tag a noun phrase as bare or modified from this corpus. The syntactic tree-patterns that I introduce in Section 2.3 (fully listed in Appendix C) are designed to match this parser's output.

Automated annotation always introduces the possibility of automated error. In Section 2.4, I discuss a sample of 500 sentences with direct object role fillers chosen at semi-random that I have further hand annotated. From this sample, we can see that some minimal errors creep in to the automatic results but that the errors should not unfairly bias results towards the Maximum Context Hypothesis or the Minimum Effort Hypothesis. Thus, from the first three sections of this chapter, it should be clear how to automatically derive my primary dependent measure (role filler modification rate) from a large-scale corpus (the BNC).

I detail the independent measures of my investigation in Section 2.5. These independent measures include control factors (like the role filler's identity) and also the three primary factors of interest that I introduced in Chapter 1: a verb/role filler's pairwise selectional strength (based on relative entropy and serving as a proxy for informativity), a verb's selectional constraint (η), and the participant role's optionality. In Chapter 3 (on direct objects), that optionality factor refers to syntactic optionality. In Chapter 4 (on instruments), that optionality factor refers to semantic optionality.

Within each of these two major studies (i.e., of direct object and instrument role fillers), I first use descriptive statistics to relate modification proportion to individual factors of interest. I then use the predictive power of linear mixed-effect models to analyze all of the independent measures in one model to see the reliability of the trends in the descriptive analysis.

Finally, in Section 2.6, I cover general statistical model properties consistent across both Chapters 3 and 4.

2.2 Modification Type: Bare vs. Modified

Some instances of role fillers are easily classified as either bare (e.g., Example 36) or modified (Example 42).

(36′) Hank ate *pizza*

(42) Hank ate *a thin-crust steak and dandelion pizza with extra mushrooms*

Unfortunately, the average naturalistic corpus sentence (in contrast with stimulus items created for experiments) is not always so black or white. For instance, if the *with*-phrase in Example 42 was instead "with soup" then most readers would not consider the post-nominal structure to be modifying the head noun. In the following section, I describe the semantic intuitions behind the labels "bare" and "modified" with respect to role fillers to draw the dividing line between categories. Many of the intuitions are based around the discussion of noun modification in Biber et al. (1999, pg. 573–656). After covering the basic intuitions behind these labels in this section, I show how they map onto syntactic frames that can be searched for in the BNC in the next section. The full set of search patterns are listed in Appendix C.

The bare category contains either bare nouns, functionally modified nouns, or compound/collocational nouns. The intuition behind the label "bare" is that the speaker could not choose a more reduced form for the role filler without omitting the role filler or making the sentence ungrammatical. For instance, the bare noun *cars* in Example 8 (repeated below) can only be reduced by full omission of the concept (as in Example 43). Examples 44 and 45, from the BNC, have the bare role fillers *boxing* and *drinks*, respectively. Further reduction in these two examples is not grammatical.

(8′) All six admitted conspiring to steal *cars*.

(43) All six admitted conspiring to steal.

(44) I've followed *boxing*, but I never heard the term 'double whammy' until it cropped up on an election poster.

(45) Maidstone ordered *drinks*.

The second class of bare role fillers have functional modifiers like determiners (as in Examples 46 and 47), possessive pronouns (as in Example 48), or demonstrative pronouns (as in Example 49).

(46) The children looked all along the bank to see if they could <u>find</u> *a boat*.

(47) Woodruffe rose and <u>led</u> *the singing*.

(48) He <u>waved</u> *his hand* expansively.

(49) Well there's a lot of people of my age that <u>wore</u> *these shoes*

I have also included what Biber et al. (1999, pg. 280–283) call semi-determiners (as in *another* from Example 50 and *no* in Example 51). Part of my justification for classifying semi-determiners as bare is that negative polarity items like *no* cannot be dropped from or replaced in an noun phrase without severely altering the meaning. Other semi-determiners like *another* do not seem to increase the relative informativity or decrease the novelty of a role filler in a way that corresponds with the Minimum Effort Hypothesis or the Maximum Context Hypothesis.

(50) I didn't want *another drink* anyway.

(51) She wore *no make-up* but had classic features, a straight nose, full lips and fine eyes.

The role filler *his neighbour* in Example 1a (repeated below) is functionally modified because it cannot be further reduced in form without becoming ungrammatical as in Example 52. Replacing *his* with a less specific modifier like the determiner *a* (Example 53) makes the sentence grammatical again with any explicit information easily recoverable from a simple inference. The inference that bridges *a neighbor* in Example 53 to *his neighbor* in Example 1a is cancellable, as in Example 54. However, this cancelling needs to be very strong and clear.

(1a′) the accused killed *his neighbour* **with a shotgun**

(52) * the accused killed *neighbour* **with a shotgun**

(53) the accused killed *a neighbour* **with a shotgun**

(54) the accused killed *a neighbour* **with a shotgun** but it wasn't his own neighbor

The final type of bare role filler is collocational forms or compounds. The term *fancy dress* (as in Example 55) is a fixed term meaning "costume". It cannot occur in the uncompounded or uncollocated form (as in Example 56).

(55) You then disappear, frighten us all to death, and suddenly, without warning, reappear, wearing *fancy dress*

(56) * You then disappear, frighten us all to death, and suddenly, without warning, reappear, wearing *dress*

These should be treated as bare mentions because they are the least marked, least modified variant for that concept. Unfortunately, these forms are not syntactically different from some of the modified variants below. Thus, these types of bare role fillers cannot be automatically differentiated from modified role fillers. I analyze a sample of 500 direct objects in Section 2.4 to verify that this inability to automate the distinction will only introduce minimal errors and not artificially bias my results towards the Minimum Effort Hypothesis or the Maximum Context Hypothesis. In Chapter 5, I propose further experiments that could reduce the number of these types of errors. To guarantee consistent annotation in my test corpus, I first needed to formalize this distinction of collocational/compound forms from truly modified forms. I will used the phrase *butcher's knife* (as in Example 5, repeated below) to help explicate this formalization. In the end, this formalization allowed me to verify that erroneous modification status tags like this would not significantly bias my results.

(5′) The court in Belfast ruled that Christie was more responsible for her actions when she killed Penny McAllister **with a sharpened butcher's knife** than was originally thought.

I have operationalized this distinction of collocational/compound forms from truly modified forms in terms of whether a new modifier can be inserted between the head noun and modifier in question. For instance, modifiers can be inserted into the left-hand noun phrases in Examples 57 and 59 but not in Examples 58 and 60.

(57) wooden bat → wooden **baseball** bat

(58) * baseball bat → baseball **heavy** bat

(59) dead bat → dead **vampire** bat

(60) * vampire bat → vampire **flying** bat

The distinction is most useful for the hand annotation task that I use to gauge error rates. In this task, I question whether *any* modifier could be inserted, not just a particular modifier. Testing for any modifier precludes false negatives due to modifier ordering rules (e.g., determiners precede adjectives).

To further complicate matters, the same surface string can referentially be either a compound noun (which I treat as bare, as in Example 61a) or a truly modified form (as in Example 61b).

(61) a. a [butcher's knife]

b. a butcher's [knife], her daughter's [knife]

In Example 61a, the head noun refers to a very particular type of knife with a certain blade shape and heft. In Example 61b, the head noun refers to a knife owned by a particular individual (e.g., a butcher or a daughter). Inserting any modifier between *butcher's* and *knife* (see Example 62 but cf. Example 63) immediately removes the semantic interpretation in line with Example 61a.

(62) a. * a [butcher's favorite knife]

b. a butcher's favorite [knife], her daughter's favorite [knife]

(63) a favorite [butcher's knife]

A modified role filler assumes that the speaker could have chosen a more reduced form or simpler construction but did not. There are five main types of modified role fillers that relate to my studies. After each description, I provide example sentences from the BNC.

Denoting modifiers (e.g., *baseball* in *baseball bat*) distinguish between concepts that are linguistically ambiguous, such as homographs and homophones (e.g., $bat_{baseball}$ and bat_{mammal}). The phrase *baseball bat* denotes the same basic level concept as $bat_{baseball}$ while blocking the concept bat_{mammal} (see Ferreira et al., 2005, for further discussion). The event descriptions in Example 64 fit these requirements.

(64) a. Larry Cummins, navigator, tells us, "I flew *the tail position*."

b. He had no hat, a lot of hair, and was not wearing *eye glasses*.

Connoting modifiers further articulate a concept. For instance, *wooden bat* is more specific than *bat*. Again using Ferreira et al.'s terms, this modification helps resolve non-linguistic ambiguity. Non-linguistic ambiguity occurs when there are two instances of the same concept in the real world (e.g., two baseball bats). Modifiers like *big* and *green* are useful for differentiating non-linguistically distinct objects (e.g., *the big/green baseball bat* and *the small/purple baseball bat*). The event descriptions in Example 65 fit these requirements.

(65) a. Because he was trying to write *deathless purple prose*.

b. I do prefer to wear *smart clothes* for work; it's still not easy for women in business to be taken seriously - I like to look the part.

c. They could not bring themselves to waste *good paper* - even if it was Fergie's."

d. He repudiated his first wife and married *a recognised Judaic princess*, thereby seeking at least a form of legal sanction.

Post-modifiers can be denotational (as in Examples 66–67) or connotational (as in Examples 68–69). I have separated post-modifiers from the above definitions because they often focus on different aspects of the head noun (mostly as a function of what English has lexicalized).[3]

(66) By this time they are approaching *the borders of the Promised Land*, and minds have turned to plans for invasion.

(67) I was then asked to give *a seminar on the inflationary universe* at Drexel University in Philadelphia."

(68) This had *the advantage of being simple to operate and relatively cheap*.

[3]Heads like *border* are particularly difficult to classify. Some may prefer to classify the prepositional phrase in this example as a complement rather than as a type of modification.

(69) Out of the corner of his eye Rincewind saw *several crossbows levelled at him.*

Some are difficult to classify clearly between denoting and connoting (as in Example 70) or contain both (as in Example 71). I used the distinction between denotational and connotational modification during my hand annotation stage in case I found significantly different modification trends between the two. Because no such differences appeared during my pilot phase, I did not analyze any such differences in my later studies.

(70) Following *a method which has been successfully piloted*, children will be grouped on the basis of individual pre-tests so that their initial explanations are either in agreement or in conflict.

(71) The crisis arose eventually over another matter, when Ken exerted successful moral pressure on one of William's Dutch courtiers to marry *the Princess's maid of honour whom he had seduced.*

Biber et al. (1999, pg. 248–257) describe several classes of "`a X of Y`" constructions that I cluster together because they are functionally equivalent for my needs. They are quantifying collectives (as in '`a group of` _____' in Example 72), quantifying nouns (as in '`a mouthful of` _____' in Example 73), unit nouns (as in '`a dish of` _____' in Example 74), and species nouns (as in '`a sort of` _____' in Example 75).

(72) Elizabeth Blackadder shows *a group of those charming still lives in which she gathers some of the objects to be found in her Edinburgh home and paints them as if they were in situ.*

(73) She looked up in surprise; she swallowed *her mouthful of chocolate digestive*, and tried to stand up.

(74) Candice is eating *a dish of beans and preserved goose.*

(75) He jumped out and came galloping down the street , waving *a sort of a club.*

Finally, event nominals (as in *the decentralization* in Example 76) seem to provide more opportunity for different types of post-modification.

(76) The two approaches disagree about how to explain *the decentralization in the 1960s and early 70s*.

Some of these event nominals are ambiguous because English allows zero-derivation of N↔V. While these example tend to be post-modified, they can be acceptably pre-modified (as with *a foxhunting ban* from Example 77).

(77) They wanted *a ban on foxhunting* - and they got it.

2.3 Automated Annotation Standards

Because I am interested in naturalistic event descriptions, I cannot rely on a pre-annotated corpus of sentences with carefully constructed verbs/role filler pairs. Further, two of my three measures of predictability (selectional strength and selectional constraint) have neither been pre-computed nor are they otherwise available for analysis without generating them myself. To address these needs, I use syntactic tree-based searches to differentiate bare arguments from modified arguments in event descriptions. These same tree-based searches can also find the appropriate role filler for the verbs of interest. These trees represent the actual annotations generated by the Charniak Parser (Charniak, 1995, 1997). I match trees using the `tgrep2` utility[4] and patterns adapted from a corpus-general structural frequency study described by Roland et al. (2007) to extract sentences containing instrument verbs.

To provide an understanding of the available tree-structures, I have recreated several of the example sentences from Chapter 1. For instance, the event descriptions in Example

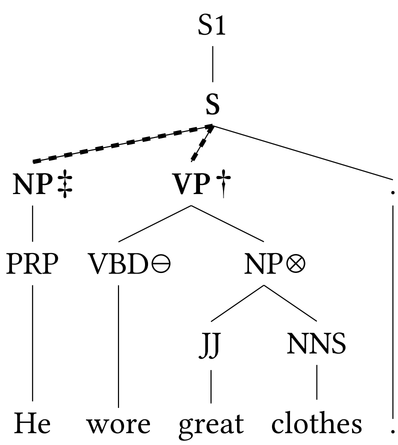
[4] `tgrep` stands for "**t**ree-based **g**lobally search a **r**egular **e**xpression and **p**rint". It can be found at: http://tedlab.mit.edu/~dr/TGrep2/

11a and Example 1a (repeated below) can be matched by looking for VP nodes (†) with an NP node sister (‡) that are immediate children of an S, S1, or Top node (three roughly equivalent variants of the sentential node).

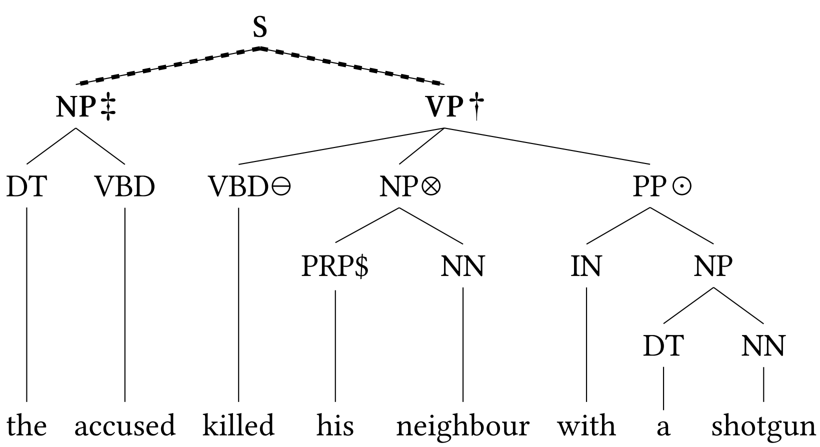
(11a′) He wore great clothes.

(1a′) the accused killed his neighbour with a shotgun

The trees for these two event descriptions are in Examples 78 and 79, respectively. I have bolded and dashed the important branches in the tree. The important nodes names are also bolded. Symbols marked beside nodes in the tree can be used to match relevant components between trees.



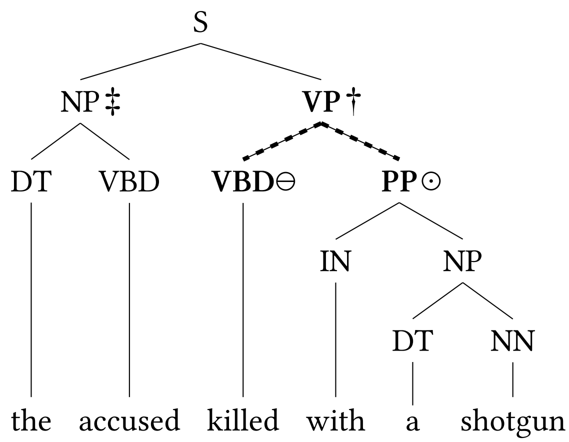
(78)



(79)

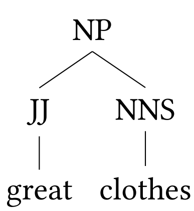
Because we are interested in event descriptions containing overt direct objects, we need to add an additional restriction so that event descriptions like those in Example 80 are not matched. The syntactically optional direct object is not present in the tree, which means it was not present in the production. As a result, we cannot automatically determine the appropriate synset for this sentence to allow us to generate a measure of role filler predictability.

(80)

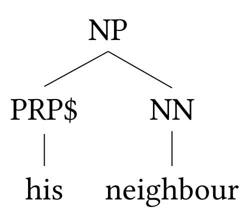


Specifically, we need to match only event descriptions for which the head verb (⊖) under the VP node has a sister NP node (⊗). This NP node (extracted in Examples 81 and 82 from Examples 78 and 79, respectively) is the role filler that gets analyzed for modification status.

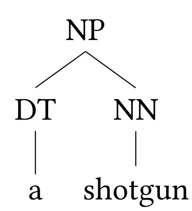
(81)



(82)



(83)



Likewise for instrument role fillers, we match the PP nodes headed by *with* (⊙) that are sisters to the head verb (⊖). Again, I extracted just this section of the tree from Example 79 in Example 83.

Example 78, of course, does not match this pattern. As a result, this event description is not returned for the instrument analysis, as it should not be.

In Appendix C, you can find the full list of `tgrep2` patterns used to match sentences for Chapters 3 and 4. Further, the important nodes of interest (e.g., the verb, the role filler, etc.) are noted in the patterns by "= *label*", like "=verb" or "=object". I used these labels as hooks for printing and extraction purposes but have left them in the patterns for clarity.

The extracted role filler is also automatically categorized (as bare or modified) by `tgrep2` patterns. For instance, the role fillers in Examples 82 and 83 are bare because the head noun is only preceded by a PRP$ (possessive pronoun) or DT (determiner), respectively. In contrast, the NP in Example 81 contains both a head noun (*clothes*) and a JJ (the adjective *great*). The full list of these patterns is also available in Appendix C.

2.4 Manual Annotations

As I mentioned above, the automated annotation process can yield erroneous tags. To ensure that these mistags do not bias the results in either direction, I hand annotated 500 event descriptions with direct object role fillers according to the actual referential choice of the speaker rather than the syntactic category determined automatically.

I randomly chose sentences across verb types to analyze. That is, rather than randomly choosing sentences (which would bias my results towards high frequency verbs), I randomly chose a sample of sentences for each verb type. This semi-random nature made sure that lower frequency verbs were also included in the error analysis. Individual verb error rates were also analyzed to guard against the possibility that high frequency verbs erred in different ways than low frequency verbs. Further, I included a slight bias towards sentences automatically tagged as modified (rather than a 50/50 bare/modified split) because there is a larger variety of ways in which a role filler can be modified than ways in which it can be left bare. The overall error rate was roughly 35%, mostly due to misparsed sentences. The errors were equally balanced between modification types. I present a more detailed breakdown of the automatic tagging errors below.

Of the 500 total sentences, 152 of them were automatically tagged as bare and 348 of them as modified (see Table 2.1).

Table 2.1: Total sentence count annotated by syntactic categorization

Type	Count
Bare	152
Modified	348
Total	500

Three event descriptions did not fit the categories described in Section 2.2. In Example 84, ‘`see a similarity between _____`’ seems like a fairly formulaic construction but it also has clearly modified sub-parts (i.e., *sarcastic southern*).

(84) So does Mark see any similarity between The Fall and these sarcastic southern pretenders?

Likewise the event description in Example 85 cannot easily be reduced to a simpler form because the definite article on *mistake* blocks the minimal form (as in Example 86).

(85) He had to think of her as Miss Crosbie so he wouldn’t again make *the mistake that had aroused his mother’s anger when he had spoken of her as Marion.*

(86) a. he wouldn’t again make *a mistake*

b. * he wouldn’t again make *the mistake*

In Example 87, the head noun is either implicit[5] or *fours* is serving as the head, depending on your preferred syntactic theory. Biber et al. (1999, pg. 519–520) include examples like *the poor* or *the Welsh* in their discussion of noun types. This construction is not an

[5]The cricket phrase ‘hit a four’ means to ‘hit a ball in such a way that you (the batter) gets four runs’. Hitting seven balls this way can be described as ‘seven fours’.

incredibly productive in English (as opposed to Spanish, for example). Within my own analysis framework, it seems most consistent to treat *fours*, *poor*, and *Welsh* as the head. Any modification to these heads would then be analyzed according to the same principles as, for instance, *car* or *pizza*. This analysis may make too strong assumptions of syntactic theory to be broadly applicable. In order to be conservative in my analysis, I treat these edge cases as errors.

(87) Glendenen, who also hit a century against Glamorgan in the NatWest Trophy , drove straight and powerfully on a slow wicket and he and Parker hit *seven fours* apiece.

Assuming similar rates in the full corpus for these types of difficult to classify role fillers, event descriptions like these should only account for 0.6% of the data.

Collocations or compounds occurred 42 times in this sample. They can occur either because the verb and direct object are tightly linked (as in Examples 88 and 89) or because the direct object is a complex noun phrase that cannot be syntactically reduced (as in Examples 90 and 91). I have to treat these as errors because they are generally frozen in form. The speaker does not have full flexibility to choose to modify or leave bare a role filler. As such, these phrases introduce additional confounds that I cannot control for.

(88) I don't want to waste *your time.*

(89) This quality of inclusion signifies the sense of the collective that women have, and from which they can draw *strength in their lives.*

(90) Satan does not realise that real freedom is found in obeying *the voice of reason.*

(91) Meredith followed *her line of sight* and saw that it rested on an oldish-looking photo.

If we extrapolate this error rate to the entire corpus, these types of errors should account for roughly 8.4% of the full corpus examples.

Finally, there were 143 event descriptions that I considered to be misparses. Misparses showed up for several reasons. First, the wrong verb could be picked out (as *wanted* in Example 92 or *printing* in Example 93).

(92) They're wanted *criminals*, you know.[6]

(93) Printing *Pictures* examines the craft of printing and processes developed in the quest for the perfect print (until mid-Mar *[sic]*).

The direct object (which I have *italicized* in examples) also could be tagged wrong. For instance, the entirety of the phrase *we move on then please* in Example 94 was tagged as the direct object.

(94) Right, I think *we move on then please*.

I also discounted phrasal verbs (e.g., *bring in* in Example 95). Phrasal verbs were frequently misparsed and thus could not be generally trusted.

(95) As we climb the stairs it sounds like people have brought *bottles* in.

Anaphoric direct objects (as in *that way* in Example 96) should have been excluded from analysis in an earlier step like other pronominal direct objects, but their part of speech tags made distinguishing them from non-pronominal, non-anaphoric direct objects difficult.

(96) He didn't think *that way* at all!

Ditransitive frames and reported speech frames were frequently misparsed (as in Examples 97 and 98, respectively). These types of errors result in the verb *say* being a highly unreliable verb type. The majority of instances of event descriptions including the verb *say* were misparses.

[6]The underlining here shows what the Charniak Parser deemed the verb. For all of the misparses cited here, my annotations reflect the given parse in the corpus, not an ideal parse.

(97) And I will show *you my city*.

(98) "What I want money can't buy," said *Big Ron*.

Misparses were by far the most common type of error. Extrapolating from the sample, these misparses will affect roughly 28.6% of all descriptions. Nonetheless, misparses seem to be balanced across verb types (except for *say* and *put*, which both had very high errors rates) and between bare and modified tags. Further, I performed all of my statistical analyses in the later chapters with and without the categorically worst performing verbs in this manual study and the results did not differ from what I have reported. Thus, we can conclude that these types of errors seem to behave as noise in the final analysis and do not unfairly bias results towards the Minimum Effort Hypothesis or the Maximum Context Hypothesis. In Chapter 5, I mention some experimental techniques that could possible be applied at a later date to further reduce this noise level.

2.5 Model Predictors

In Sections 2.2 and 2.3, I defined my dependent variable (modification status) and covered the means to automatically determine it from a syntactic tree-parse. The Maximum Content Hypothesis and the Minimum Effort Hypothesis each make claims about the relationship between predictability and modification status. In order to compare the relative dominance of the production strategies underlying these hypotheses, I need robust and automatable metrics for predictability. A robust measures means that it will be reliable despite noisy input. An automatable measure allows me to analyze a corpus far larger than would be reasonable to hand-annotate.

In this section, I define and explain how to determine these independent measures to set me up to analyze them in Chapters 3 and 4. I will first define the control measures. Using control measures will underscore that my final results are not underlyingly caused by well-attested and simpler phenomenon or properties. These include factors like the role filler's identity. The identity of the role filler could influence the speaker's choice of modification in a way outside of the interest of my study on informativity. For instance, *wine* can refer to a color or an alcoholic beverage. The number, type, and variety or things said to modify a color term may very well be different from those of an alcoholic beverage. As such, there is presumably an inherent difference between the baseline modification rates of color terms and alcoholic beverages. We need to include control factors like this in the study to create the best defined baseline for the speaker's choice.

The control factors help control and explain some of the noise in the data that is underlyingly unrelated to the factors of interest. Because statistical models are greedy in their explanatory or descriptive power, they attempt to account for the maximal data possible. This greediness can result in a high-level factor (like argument predictability) being attributed as the underlying cause of something better explained by a low-level factor like lemma or part-of-speech. This would be equivalent to trying to predict pregnancy rates at the individual level (e.g., 'can this individual get pregnant') and using the individual's name but not sex as a predictor. You will find that the name *Grace* is highly predictive of can-get-pregnant while *Fritz* is highly predictive of cannot-get-pregnant. The real underlying trend is that only women can get pregnant and *Grace* is more likely to be a woman's name than *Fritz*.

After creating the baseline models containing only control factors, I introduce my three factors of interest (argument predictability, a verb's η or selectional constraint, and the verb's optionality class) in a complete model. Does the complete model do a better job of predicting our speakers' choices than the baseline model? If so (and it will be so), then introducing my factors of interest significantly improves our understanding of speakers' choices.

2.5.1 Baseline Independent Measures

The first lexical property that I use as control factor needs to capture a word's baseline activation level. Intuitively, words with a high baseline activation level should be generally more accessible and generally easier to name because speakers always need access to them. Frequency is often used a proxy measure for baseline activation level. However, relying solely on frequency can lead us astray in many contexts. Word frequency tracks the total (or expected) number of times a word occurs in normal speech. Contextual diversity tracks the number of different contexts (or documents) a word occurs in.

Contextually diverse words should be generally more accessible and generally easier to name because speakers always need access to them. In contrast, high frequency does not entail that a word is likely to be accessed at any given time. Thus, contextual diversity should serve as a better proxy measure for the baseline activation and accessibility of a lemma[7] than frequency. Adelman et al. (2006) find exactly these trends when comparing frequency to contextual diversity as a predictor of word naming and lexical decision times. Predictive models with contextual diversity do better than predictive models with just frequency. Moreover, adding contextual diversity to a model that already includes frequency improves the model's fit, whereas the opposite is not true. This unidirectional improvement shows that contextual diversity adds information above and beyond frequency. My values for contextual diversity are taken from the SUBTLEX Corpus of TV and movie subtitles (van Heuven et al., 2013). Each TV episode or movie that a word occurs in increases that word's context count by one, regardless of how many times it occurred in said TV episode or movie.

In terms of a concrete example, words like *look*, *only*, and *over* have relatively low frequencies in comparison with other words of equivalent contextual diversity. In contrast,

[7]I use the term *lemma* here to mean roughly citation form, as is standard in computer science.

words like *slipstream*, *Thatcher*, and *Madeleine* have relatively high frequencies for their contextual diversity values. In a particular context, these words may occur frequently but are not likely to occur in any given context. Words in the first set are evenly spread across contexts, topics, or documents and are a good bet to show up at any time. Using frequency as a metric for words like *look*, *only*, and *over* would result in speakers regularly *under*predicting their occurrence because frequency only informs us as to how heavy-handed the use of the word is. For comparison, salt and pepper are fairly ubiquitous flavors. Although both flavors are used equally across the board, salt tends to be used in higher concentrations than pepper. Analyzing the total amount of salt and pepper used over a year of cooking may incline someone to assume that pepper is used rarely. In actuality, pepper is used commonly but sparingly.

Words in the second set clump together and are only a good bet to show up if they have already shown up in the current context. Using frequency as a metric for these words would result in speakers regularly *over*predicting their occurrence because speakers will assume that a highly frequent term is equally likely to occur in any given context rather than it being unlikely is most contexts and incredibly like in a minority of contexts. Returning to a food analogy, hops are a fairly rare ingredient. They are almost exclusively used in the context of brewing beer and in this context are used in large quantities. We should not expect hops to show up in any given recipe but any recipe including hops will likely include lots of hops.

Within my investigation into speakers' modification choices, it seems reasonable to assume that highly contextually diverse role fillers (e.g., *room* with a contextual diversity of 74.52 out of 100 according to van Heuven et al., 2013) would be modified at a different rate than less contextually diverse role fillers (e.g., *cell* with a contextual diversity of 17.39). Fillers like *room* could require more contextual disambiguation (often in the form of modification) than fillers like *cell*. From simple inspection by the author, *room*-like fillers are often more generic and/or bleached terms than *cell*-like terms. This genericness could also impact modification rates. Likewise, highly contextually diverse verbs (e.g., *kill* with a contextual diversity of 67.95) could influence the role filler realization in a different manner than less contextually diverse verbs (e.g., *murder* with a contextual diversity of 24.39).[8]

The second lexical property that will serve as a control factor is the underlying identity of the role filler's head noun. It seems reasonable to assume that certain lexical items have individual biases with respect to how and how often they are likely to be modified. Ideally, we would associate each head noun with the speaker's intended underlying concept. The concept itself is no doubt the larger determiner of how and how often a head noun can be modified. Unfortunately, no such function from string representation to underlying concept exists.

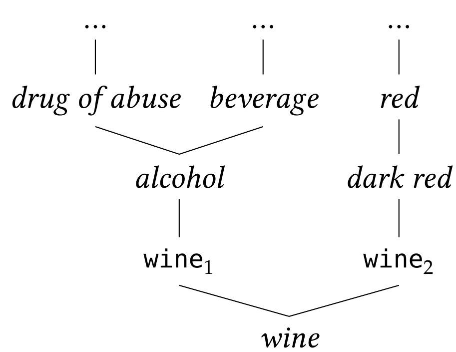
Functions do exist to map from string to citation form. This process is the first step in the right direction. It allows us to make the appropriate generalization that *wine* and *wines* should behave similarly. However, we would still miss the important generalization that *wine* and *beer* may also behave quite similarly. I want to associate a head noun with the concept it *could* represent since I cannot automatically associate it with the exact concept it *does* represent. WordNet allows me to make these connections because it is a linked network of lexical information containing sub-networks by part of speech. Hence, following Resnik (1993; 1996; see Abe & Li, 1996; Ciaramita & Johnson, 2000; Clark & Weir, 2001, for similar approaches) I replace the lemma representation with a WordNet synset (Miller, 1995; Fellbaum, 1998). The WordNet synset (which I explain below) serves as a placeholder for the conceptual abstraction the speaker intends. Using a synset rather than the lemma reduces data sparsity issues and improves the semantic subcategorization

[8] The pair of terms *kill* and *murder* share a special contextual diversity relationship because the former is a strict superset of the latter. Every *murder* event is a *kill* event but not the other way around.

generalizations that the system can be attuned to, essentially getting us as close to the underlying concept as we can reliably do automatically and on a large scale. First, I will explain what a WordNet synset is and then I will provide details on how their use helps solve the two issues just mentioned.

I exclusively use hypernym/hyponym and synonym links within WordNet's noun taxonomy for my studies. Hyponymy encodes class membership between concepts. For instance, *wine* has a drink-related sense and a color-related sense. These are encoded as ($\texttt{wine}_1$ `IS-A` *alcohol*) and ($\texttt{wine}_2$ `IS-A` *dark red*), as in Figure 2.1. Synonymy encodes similar meanings between different word forms (e.g., the synset for $\texttt{wine}_1$ has lemma representations of *wine* and *vino*). Thus, the different word forms *wine* and *vino* have equivalent meanings that all map to the same concept called `wine.n.01`.[9] That concept is a member of the concept `beverage.n.01`, which can be expressed with the roughly synonymous word forms of *beverage*, *drink*, *drinkable*, *potable*. The sets of synonyms are called synsets.

Figure 2.1: Beverage, color, and wine tree



Miller and colleagues intentionally designed WordNet to be a differential theory of meaning rather than a constructive theory of meaning. That is, as a differential theory of meaning, the taxonomy does not provide enough information to uniquely construct

[9]Synset labels, synset word forms, and hypernym/hyponym relations are take from WordNet v. 3.1, the same version that all my analyses are derived from.

the concept but it does provide enough information to identify the concept from a set of known concepts. A user can figure out the real world concept intended by the WordNet concept by inspecting its hypernyms, hyponyms, and synset word forms. The conceptual definitions are conveyed through the implicit inferences encoded in the type hierarchy.

Further, WordNet is not a simple tree, much like our actual mental lexicon is not a simple tree of connections. A complex tree with multiple inheritance and polysemous word forms allows us to encode richer relations and more types of dependencies. For instance, one concept can be a member of multiple disjoint classes (e.g., `alcohol.n.01` is a `drug_of_abuse.n.01` and `beverage.n.01`). Additionally, one word form can be associated with multiple synsets (e.g., *wine* is associated with `alcohol.n.01` and `dark_red.n.01`). When we see the word *wine*, it should be treated as possibly activating all of its synsets and superclass synsets.

Each of these conceptual synsets inherits properties from all of its superclasses (or hypernyms). Hence, a statement such as in Example 99 could be, in terms of which concepts are activated but not necessarily activation strength, equivalent to any of those in Example 100. This extensional approach to interpretation is what allows the use of synsets to capture subcategorization generalizations. The most common event descriptions of a *drink*ing event converge on the concept `fluid.n.01` rather than the more specific concepts of `wine.n.01` and `coffee.n.01` (the drink, as opposed to the tree, color, or bean) or the less specific concepts of `physical_entity.n.01` and `matter.n.03`. This convergence on a synset can be thought of as a process of consensus building across examples. Based on two examples with the same verb, we can choose a single synset that sits in the conceptual hierarchy of both fillers. When we add a third example, we will likely need to slightly adjust the best synset. Starting with a different set of two example sentences could begin the process in a different part of the synset hierarchy. Given enough examples, though, our preferred synset (or synsets) will converge to the optimal answer no matter what our initial examples were.

(99) I drank some wine.

(100) I drank some...

a. `wine.n.01`
b. `alcohol.n.01`
c. `drug_of_abuse.n.01`
d. `beverage.n.01`
e. ...
f. `wine.n.02`
g. `dark_red.n.01`
h. `red.n.01`
i. ...

In the same light, backing off from the lemma to the best convergent concept reduces data sparsity problems. Table 2.2 shows a sampling of direct object role fillers in the BNC. The left-hand side show the frequencies and expected rate of occurrences for these role fillers when used in *drink* event descriptions. Based strictly on these numbers, our expectation of ever seeing a role filler like *gin* or *juice* is zero because these specific combinations never occur in the BNC. Based on our real world knowledge, those expectations are clearly the wrong conclusion. The right-hand side of Table 2.2 shows the convergent synset concepts and those frequencies and expected rate of occurrences. Based on this table, our expectation for *gin* and *juice* are very high (65%), which relative to something like the expectation for *stuff* (`substance.n.01` at 1%), seems much more reasonable.

A full example for determining the synset at the convergence point is worked out in Appendix B. This example should help provide strong intuitions about what terms and trends bubble up to the top. In simple terms, I start with the word lemma of the role filler's head. I use WordNet to split this lemma into each of its synsets and each of those synsets' hypernym synsets, all the way up the taxonomy. The word *wine* is replaced with the set (`wine.n.01`, `alcohol.n.01`, ...). The word *stuff* is replaced with the set (`stuff.n.02`,

Table 2.2: Direct Object Role Filler Frequencies for the Verb *drink*

Lemma Count	Lemma	Expected Lemma Occurrences	Convergent Synset	Synset Count	Expected Synset Count
92	coffee	14%	`fluid.n.01`	419	65%
89	tea	14%	`fluid.n.01`	419	65%
44	water	7%	`fluid.n.01`	419	65%
29	wine	4%	`fluid.n.01`	419	65%
23	alcohol	4%	`fluid.n.01`	419	65%
11	drink	2%	`fluid.n.01`	419	65%
8	stuff	1%	`substance.n.01`	21	1%
1	chablis	<1%	`fluid.n.01`	419	65%
1	beaujolais	<1%	`fluid.n.01`	419	65%

`material.n.01`, ...). Then, for each synset in turn, I calculate the relative entropy of the verb+synset with respect to all uses of the synset (as in Equation 2.1).

$$\begin{aligned} &p(\,\texttt{wine.n.01} \mid drink\,) * log \frac{p(\,\texttt{wine.n.01} \mid drink\,)}{p(\,\texttt{wine.n.01}\,)} \\ &p(\,\texttt{alcohol.n.01} \mid drink\,) * log \frac{p(\,\texttt{alcohol.n.01} \mid drink\,)}{p(\,\texttt{alcohol.n.01}\,)} \\ &p(\,\texttt{dark\_red.n.01} \mid drink\,) * log \frac{p(\,\texttt{dark\_red.n.01} \mid drink\,)}{p(\,\texttt{dark\_red.n.01}\,)} \\ &\ldots \\ &p(\,\texttt{matter.n.01} \mid drink\,) * log \frac{p(\,\texttt{matter.n.01} \mid drink\,)}{p(\,\texttt{matter.n.01}\,)} \end{aligned} \tag{2.1}$$

The synset that generates the highest relative entropy value is the convergence point. That synset is treated as the target conceptual representation used as a control factor.

2.5.2 Independent Measures of Interest

In the end, we also need measures of specific verb/role filler predictability, general verb predictability, and filler optionality. These three measures should help provide a clearer understanding of the relationship between informativity and production choices. The value of the relative entropy (discussed in the previous section) for a synset/verb pair serves as the first of my factors of interest. This value represents the selectional strength between the verb and role filler. A high value (like between *drink* and `fluid.n.01`) indicates that words conceptually related to *fluid* (e.g., *drink*, *beer*, *coffee*, *wine*, etc.) are more strongly associated with *drink*ing event types than with an unknown, arbitrary event type. In contrast, a low selectional strength with the verb *drink* occurs because synsets like `(coffee)` are too specific or synsets like `matter.n.03` are too generic.

The aggregate selectional strength (called η or selectional constraint) for a verb is the second factor of interest. For every occurrence of a verb, we calculate the maximum converging selectional strength with its role filler The sum of all of these selectional strengths is mathematically the η value and conceptually the selectional constraint. High η values mean a verb is most often paired with a highly predicted role filler. Low η values mean that there are not many role filler types that share converging synset representations and, hence, the role fillers come from a more disparate set of concepts.

The final factor of interest is the verb's categorization with respect to the optionality of the role filler. In Chapter 3 (where I investigate verb/direct object role filler pairs), this categorization is the split between syntactically optional and syntactically obligatory direct objects. In Chapter 4 (where I investigate verb/instrument role filler pairs), this categorization is the split between semantically optional and semantically obligatory instruments. See Sections 1.1.3 and 1.1.4 for further discussion of these categories. Each verb's category is listed in the table in Appendix A.

2.6 Model Overviews

My primary inferential statistics in Chapters 3 and 4 are derived from linear mixed-effect models. The mixture of effects comes from the combination of fixed effect predictors and random effect predictors. The intercept of these regression models reflects the baseline, expected result without taking into account any of the predictors. For my studies below, this intercept is the average modification rate across all role fillers. Random effects can be thought of as idiosyncratic variation conditioned on the predictor. For instance, different role filler's head synset will have different baseline modification rates at the very least because of differences in the underlying semantics of the synsets. Using the synset as a random effect allows that baseline to vary slightly based on the role filler without affecting the rest of the model. Experimental participant is frequently used as a random effect in reading time studies to accommodate the naturally different baseline reading rates individuals have unrelated to the study in question.

Fixed effect coefficients (or β-values) will be the more interesting value to analyze. These coefficients are the same as those in a traditional regression model. Positive values for my studies mean that increases in the predictor value increase the likelihood of the role filler being modified. Negative values mean that increases in the predictor value decrease the likelihood of the role filler being modified. I assume an alpha level of 0.05 for all tests.

Not all of my collected data meet the requirements to be used in a linear mixed-effect model. Analyzing this data will be helpful for understanding the general trends across the corpus with respect to my measures of interest. For instance, my paired population comparisons will all be non-parametric comparisons. Neither selectional constraint (η) for verbs nor selectional strength values for verb/role filler pairs can be assumed to have normal distributions, thus failing a requirement for using parametric tests. Further, analyses of selectional strength include repeated sampling of verb/role filler pairs when the same pairs occur in multiple event descriptions. As such, I use a Wilcoxon signed-rank test. This test is similar to a *t*-test, in that it is designed to assess if two samples are from the same or different populations, but is licensed for non-parametric, repeated measures.

Second, I calculate effect size (specifically, Cohen's *d*; Cohen, 1988) for population comparisons. I need to be able to compare results between direct object and instrument constructions, for which I have vastly different sample sizes. Since effect size reports the magnitude of a relationship, I can make meaningful comparisons between these two constructions. Effect sizes are usually binned into categories (small, medium, and large) to provide intuitive labels for the magnitude of the expected difference between populations, if it exists. I will report those bins in addition to the actual values so that the reader can choose their preferred means of interpretation (i.e., categorical bin or numerical Cohen's *d*). Pitch differences between adult men and women qualify as a large effect. To ground the lower end of the categorical bins where some of my effects occur, I will refer to sample linguistic data compiled by Newman et al. (2008) in discussing gender-based differences in language use. They used a basic word count strategy (see Pennebaker et al., 2001) to compare sentences sampled from a range of sources and genres to look for systematic and contextually determined differences between female and male language patterns. The distribution of swear words had an effect size of $d = 0.22$, with a bias towards men preferring to use them. Words related to the home and words relates to sports both had an effect size of $d = 0.15$, with the former preferentially used by women and the latter by men. The difference in emotion word usage had a very small effect size of $d = 0.11$, biased towards women. It is important to note that anger words and swear words were two categories of terms that full under this umbrella of emotion words, but were both strongly biased towards usage by men. Critically, effect size is not an inferential statistic, so it does not tell us whether or not any perceived differences are reliable. For that, we use the Wilcoxon signed-rank test.

I began the chapter formalizing my dependent measure (argument modification rate). Comparing the contexts of bare and modified role fillers give us insight into speakers' choices and the forces influencing their choices. I described the referential intuitions behind the structural patterns that I use to pull example sentences from the BNC. These example sentences serve as the body of evidence I will analyze using statistical models in Chapters 3 and 4.

At the end of this chapter, I defined the independent measures (predictors) used to characterize and quantify the influences behind speakers' choices. These predictors are a combination of control measures (to gauge an accurate baseline model that assumes my factors of interest are not present or available) and measures of interest (that quantify informativity for the speaker).

As a reminder, I am interested in the extent to which argument predictability affects the descriptive choices of a speaker. On the one hand, speakers could use their internal expectations of informativity to help them clarify those parts of an event description that may be most surprising. I have caricatured these individuals as solicitous speakers maximizing the context of all event description components. On the other hand, speakers could use their internal expectations to allow them to maximize informativity. I have caricatured these individuals as rational agents minimizing the effort of producing the event description.

In Chapters 3 and 4, I use event descriptions from the BNC to show that production choices in line with the Maximum Context Hypothesis tends to dominate production choices in line with the Minimum Effort Hypothesis. Speakers' normal choices are better modeled by the solicitous speaker than the rational agent speaker. That is, speakers are biased by their internal knowledge of argument informativity to modify less predicted role fillers more often than highly predicted role fillers. This bias holds true when looking as the predictability of individual verb/role filler pairs and when accounting for the generalized predictability of all of a verb's role fillers.

Further, I am able to use this corpus of event descriptions to uncover general trends in participant role predictability as a function of its strength or reliability in a verb's mental lexicon entry. Roles for a verb specified to be syntactically optional are more predictable than those specified to be syntactically obligatory (see Resnik, 1993, for similar quantitative results). Roles specified to be semantically obligatory are more predictable than those specified to be semantically optional (see Koenig et al., 2003; Yun et al., 2006, for similar qualitative results). As explanation for why these verbs classes trend together, recall that semantically obligatory roles and syntactically optional roles both require strong connections between the verb and likely fillers of the role.

Finally, in Chapter 5, I explain the implications of my results for computational models of semantic relatedness and psycholinguistic models of processing difficulty.

Chapter 3

Comparing Syntactically Obligatory and Optional Roles

DO or DO not.

— Yoda, *The Empire Strikes Back*

3.1 Overview

In this chapter, I use verb/direct object role filler pairs to investigate the Maximum Context Hypothesis and the Minimum Effort Hypothesis. Specifically, I will track the relationship between direct object role filler modification and three predictors of modification: the syntactic optionality of the direct object (given the verb), the verb's selectional constraint (η) with respect to all its direct object role fillers, and a verb/direct object role filler pairwise selectional strength. All three predictors of modification provide insight into the informativity or unexpectedness of an event description unfolding in the particular manner that it does. I will use these indicators of predictability to show that speakers tend to abide by the Maximum Context Hypothesis more often than the Minimum Effort Hypothesis. Namely, a speaker is more likely to provide additional details about a direct object if that direct object represents a less common role filler for the given event. This tendency to provide additional details is even stronger when the verb has weak selectional constraints.

The relationship between the syntactic optionality of the direct object participant role and the predictability of that role's fillers is an important underlying premise. For one, Resnik (1993) argues that one of the underlying mechanism for this lexical switch (of a verb licensing syntactic optionality) is partially based on the predictability of the direct objects associated with it. Licensing syntactic optionality is a categorical distinction rather than a cline of more and less omissible.[1] A necessary but not sufficient condition for a verb to license direct object omission is having highly predictable direct object role fillers that fit into a clear super category[2]. Levin (1993), in discussing direct object omission, claims that the implicit assumption of omission is that the most frequent direct object is to fill that particular role. One of the intuitions that Resnik (1993) tested in his studies was that syntactically optional verbs themselves encode some of the features of their highly predictable direct object role fillers. Thus, accessing the verb alone (without an explicit direct object) is sufficient (and reliable) to activate both the verb's concept and the most likely direct object filler concept (or its features)[3].

I am interested in this intuition in as much as it concretely specifies stronger mental lexicon connections between certain verbs and their direct object fillers and quantifies the relative entropy between specific verb and direct object participant role filler pairs.

[1] This cline should not be confused with Lehrer (1970)'s distinction between Type 1, 2, 3, and 4 verbs, or between indefinite object and specified object alternation

[2] Recall in Section 2.5 where I discuss how to calculate the converging synset for an argument. This convergent synset is a super category of all its descendent synsets or hyponyms. More generally, a super category is any hypernym. This term can also be applied to a shared hypernym between any two or more words or concepts.

[3] I do not distinguish between the possibilities that the verb is activating the direct object filler concept, features of the concept, or some super category that uniquely picks out the likely candidates. My results and analyses remain the same regardless of these specifics.

Syntactically obligatory verbs should have, on average, weaker mental lexicon connections and higher entropy (i.e., be less informative or mutually predictive) with respect to their direct objects than syntactically optional verbs. By definition, verbs with weaker selectional constraints (low η values) will have weaker mental lexicon connections to role fillers and less predictable role fillers with respect to their direct objects than verbs with stronger selectional constraints (high η values). Are speakers (or writers) sensitive to these different levels of informativity and mental lexicon connectedness in their productions?

In order to verify an underlying premise of my model, I will replicate Resnik (1993)'s studies that show syntactically optional direct object verb's have significantly higher η values than syntactically obligatory verbs. The premise I need to verify is that syntactically obligatory verbs, as a category, have less predictable role fillers (and, as a consequence, weaker connections to those roles and fillers) than syntactically optional verbs. I then extending Resnik's original studies by using a larger corpus and by using a larger verb set in order to show that his results (and this premise) are not just an artifact of a small corpus or of high frequency verbs. I predict that this relationship between the average η value (a measure of selectional constraint) for syntactically obligatory and syntactically optional verbs will only hold at the verb category level. Namely, some individual syntactically obligatory verbs may have higher η values than some individual syntactically optional verbs. The average selectional constraint across all syntactically obligatory verbs, however, will be weaker than the average for syntactically optional verbs.

According the Minimum Effort Hypothesis, speakers should opt to make productions maximally informative without being too verbose. Highly predictable direct objects should be modified (in order to increase their informativity) or be omitted (in order to decrease the unnecessary verbosity). According the Maximum Context Hypothesis, highly predictable verb and direct object pairs should have the strongest activation weights for their links, which should lower the threshold required for co-activation. Less predictable pairs will require more effort, which in turn makes them prime targets for modification.

Ideally, I would be able to use direct object inclusion or omission as my dependent variable. Speakers who believe the direct object to be completely recoverable (and, thus, fully uninformative) would reasonably omit any role filler. Unfortunately, syntactically obligatory direct object verbs, by definition, cannot be used in such a study. In these cases, the speaker lacks the option to omit the role filler. Instead, I will compare modification rates of direct object participant role fillers when a head noun is already present. The first step in this process, with the help of corpus analysis tools, is to automatically recognize event descriptions that contain a direct object noun. I can then syntactically analyze these nouns or noun phrases (depending on the construction) with respect to their modification status. Is the head noun bare or is it modified in any way? The underlying assumption of modification as a dependent variable is that a modified direct object is more informative than the same bare direct object would be for describing a particular event. Do people tend to choose to produce the more informational modified direct object when that direct object is already highly predicted by the verb (the Minimum Effort Hypothesis) or do they choose to produce the bare direct object because activation is so easy (the Maximum Context Hypothesis)?

In Section 3.1.1, I start out by explaining η and providing examples of η values for different verbs and the types of verb/direct object pairs that result in high or low selectional strength measures. In Section 3.2, I replicate Resnik's studies that compare η values between verb classes and extend those studies to larger datasets by verb type and sentence token. In Section 3.3, I analyze simple predictors of modification proportion to showcase the basic trends between modification proportion, verb class, selectional constraint (as measured by η), and selectional strength. In Section 3.4, I will look at modification proportion as modulated by informativity using linear models to look at the individual impacts of the factors and their interactions.

3.1.1 Direct Object Selectional Constraint and Selectional Strength

Selectional constraint and selectional strength are deeply tied to each other. Verbs have strong selectional constraints because they share high selectional strength with their average role filler. It is important to remember that these verbs do still have low selectional strength with some of their role fillers. Likewise, verbs with weak selectional constraints can also sometimes have high selectional strength with some of their role fillers.

To provide a better understanding of the range of selectional strengths for verbs and how they interact with selectional constraint, I have chosen four representative verbs of high and low η values with several respective direct objects of high and low selectional strength for these verbs. These verbs and their related direct objects are in Table 3.1. The η value for each verb ($0 <= \eta <= 1$, $M = 0.33$, $SD = 0.17$) is in parentheses. The verbs *bring* and *enter* are on the lower end of the range of η values. The verbs *drink* and *open* are on the higher end. Of the four verbs, *drink* is the only syntactically optional direct object verb.

To better understand what makes a weak selectional constraint verb have a low η value, let us look closer at the general details of verbs like *bring* and *enter*. Intuitively, both verbs conceptually allow for a wide range of direct objects and syntactically require a direct object. These are both good indicators that the verb is likely to have weak selectional constraints. Looking at specific properties for *enter*, *room*-like direct objects[4] have a particularly high selectional strength. These *room*-like direct objects are both very frequent with *enter* (the left-hand side of Equation 3.2) and relatively infrequent as direct object of other verbs (the right-hand side). However, a small number of high selectional

[4] As covered in Chapter 2, I do not use the direct objects themselves in these measures. Instead, I use the WordNet synset (out of all synsets associated with a noun) that has the highest selectional strength with the verb.

Table 3.1: High and Low η Verbs with High and Low Selectional Strength Direct Objects

Low Select. Str.	Verb	High Select. Str.
	enter (0.12)	
window		room
wood		
	bring (0.25)	
whisky		flower
box		water
	open (0.68)	
letter		door
window		eye
	drink (0.88)	
blood		coffee
stuff		alcohol

strength direct objects like *room* for *enter* are not enough to guarantee a strong overall selectional constraint. Relatively more of *enter*'s direct object role fillers have low selectional strength with the verb. These low selectional strength direct objects (e.g., *window* and *wood*[5]) are either infrequent as direct objects of *enter* (the left-hand side of Equation 3.2), frequent as direct objects of other verbs (the right-hand side), or both.

The selectional strength of these verb/direct object pairs is a composite factor that integrates both the co-occurrence of the verb/direct object pair and co-occurrence of the direct object with other verbs. I have repeated the formula for selectional strength in Equation 3.1 as derived from Equation 1.8, the formula for a verb's η. Plugging in actual corpus frequencies in Equation 3.2 and Equation 3.3 would provide the numerical backing for these relative frequency intuitions I described in the previous paragraph.

[5]I do not distinguish between the senses of $enter_{room}$ and $enter_{contest}$. The direct object *contest* is actually near the higher end of the selectional strength range for the verb *enter*. As such, the selectional strength measure can be said to allow multiple sense to co-exist within the same set of measures. The η value, however, will be strongly biased by both the most selective and the most frequent of the different senses.

$$p(\, c \mid v\,)\, log\left(\frac{p(\, c \mid v\,)}{p(\, c\,)}\right) \tag{3.1}$$

$$p(\, room \mid enter\,)\, log\left(\frac{p(\, room \mid enter\,)}{p(\, room\,)}\right) \tag{3.2}$$

$$p(\, wood \mid enter\,)\, log\left(\frac{p(\, wood \mid enter\,)}{p(\, wood\,)}\right) \tag{3.3}$$

At the other end of the η scale, verbs like *open* and *drink* have a much narrower range of direct objects. These types of verbs are good for understanding why a verb can end up with strong selectional constraint (a high η value). Intuitively, the typical direct objects of a verb like *drink* fall into a smaller set than the direct objects of a verb like *enter* (or even *open*). As a reminder to the reader, this smaller set of direct objects is really a smaller set of convergent WordNet synsets. I have already abstracted away from the terms *room* (e.g., in *enter a room*) and *beer* (e.g., in *drink a beer*) to the most appropriate synset. While *enter*-like verbs tend to have their direct object pairs dominated by weaker role fillers with which they share lower selectional strength, *drink*-like verbs tend to have more direct object pairs with which they share higher selectional strength.

In Section 3.2, I will explicitly test the intuition that direct objects of syntactically optional verbs have higher η values. For now, it is more important to note that, while *drink* is syntactically optional, the verb *open* is syntactically obligatory. A high η value cannot be the sole determiner of optionality.[6] If this were the case, we should be able to sort all verbs by their η value and set a threshold that exactly separates the syntactically optional verbs from the syntactically obligatory verbs. The lessened diversity of direct objects for

[6]To the best of my knowledge no one has proposed a testable explanation of why a high η value is not a sufficient condition to determine optionality. One possibility is that it may be difficult for speakers to distinguish contextually predictable from globally predictable. A high η implies globally predictable role fillers but does not determine the specific contextual predictability. As such, requiring other conditions be met before an argument can be licensed as optional could serve as a safety catch. It conservatively prevents over-elision or over-reduction until some other requirement(s) is/are also met.

high η verbs can also be observed when comparing the high and low selectional strength direct objects for *drink*. A natural class that encompasses low selectional strength direct objects for *drink* like *blood* and high selectional strength direct objects like *coffee* is easy to generate and fairly specific: potable liquids. In contrast for the verb *bring*, a natural super category for *whisky* and *flower* is much harder to extrapolate.

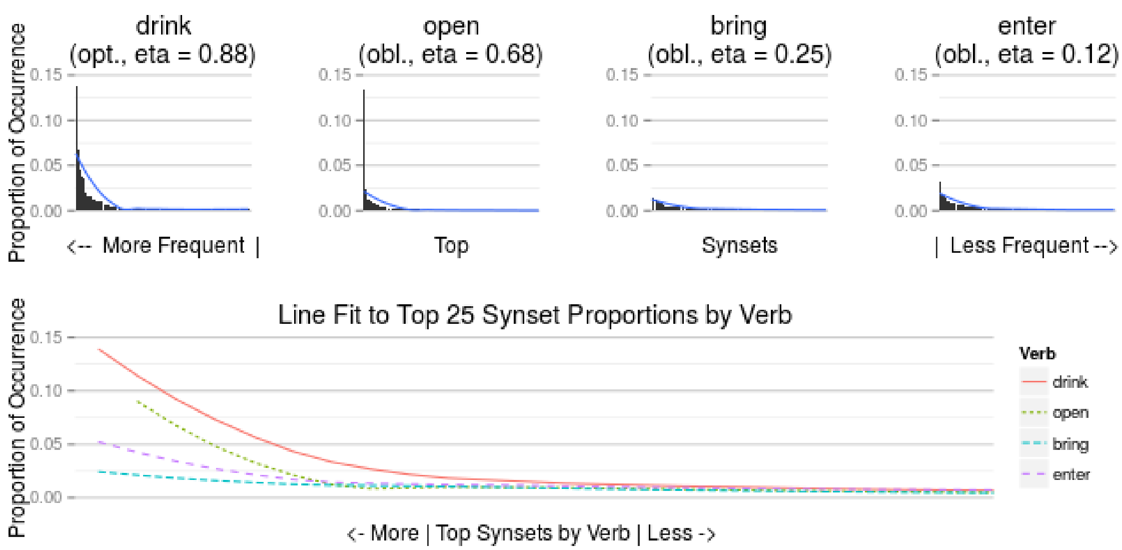


Figure 3.1: Proportion of Occurrence for Direct Object Synsets with Sample Verb

Figure 3.1 makes the intuitions described above more concrete. The x-axis represents the most frequent WordNet synsets used as direct object role fillers for each verb. The most frequent synsets for a verb are on the left and less frequent synsets are on the right. Each bar is a direct object concept. In the top graphs, the bar height represents the proportion of occurrences that the given synset accounts for with respect to the given verb. Looking at the leftmost bars, we can determine the proportion of event descriptions for the given verb that evoke the most frequent synset. The vertical bars at each x-axis tick decrease in length as we move right along the graph, because each subsequent synset is the convergence point of a smaller and smaller proportion of the event descriptions. The most common synset for *drink* occurs in roughly 15% of all event descriptions. In contrast, the most common synset for *bring* only occurs in 3–4% of all *bring* event descriptions.

The bottom graph combines the smoothed trend line for each of the four verbs to facilitate direct comparison of mention proportions. In this graph, it is clear that synset proportions for *drink* start off higher than for *open* and appreciably higher than for either *bring* or *enter*. In more technical terms, the majority of the probability mass for the high η verbs is concentrated on the left side of the graph (and associated with a smaller number of distinct synsets) while the probability mass for the low η verbs is more evenly spread across the graph (and associated with a larger number of distinct synsets).

The high η values of *drink* and *open* are achieved through a small number of very frequent concepts. The verb *open* has three synsets that account for roughly 47% of the total direct object mentions. The verb *drink* has three synsets that account for roughly 35% of the total direct object mentions. In comparison, *bring* and *enter* are associated infrequently with a larger number of concepts. The top three synsets for either of the low η verb is only going to account for 13% (*enter*) to 6% (*bring*) of the total occurrences.

3.2 Selectional Preference and Implicit Objects

This section is intended to underscore the relationship between a verb's selectional constraints and the quantity of information it carries about its arguments. Strong selectional constraints require that the verb encode lots of information about possible arguments. Weak selectional constraints mean the verb can be more agnostic to their possible arguments. Resnik (1993, pg. 1) demonstrates selectional constraints "can be expressed as an information-theoretic relationship" through several experiments replicated below. He (among others like Lehrer, 1970; Fillmore, 1986; Rice, 1988; Fellbaum & Kegl, 1989) has focused on the connection between direct object omission and predictability. This dissertation is primarily concerned with how participant roles are realized, as opposed to when they are omitted.

Resnik's thesis, as born out by experiments, is that those classes of verbs which license direct object omission will have higher average selectional strength between each verb and its role fillers than will those verbs which always require an overt direct object.[7] Resnik (1993, pg. 84) distinguished the two verb classes according to the following criteria:

> Each verb was classified as object-drop or non-object-drop. A verb was classified as object-drop only if (a) some sense of the verb is annotated with both V and V+O in (Sinclair, 1987), and (b) that sense is "close enough" to the central meaning of the verb, as opposed to an extremely specialized sense. The latter criterion is a question of personal judgment: some sense of each verb in (101) and (102) permits both subcategorizations, but in cases like (102) [Resnik] decided the senses permitting the alternation were too specialized to warrant categorizing the verbs as object-drop.

(101) a. John called (someone) at 3pm.

b. John packed (a suitcase) quickly before leaving.

c. John stole (some money) and was caught.

(102) a. John opened (a discussion) with a question.

b. John showed (a work of art) in New York.

c. The missile hit (a target) and exploded.

Lexicon 3.1 lists all the verbs used in Resnik (1993)'s Brown Corpus studies on selectional constraints and object omission. I have re-labeled the groups according to the syntactic

[7] As Resnik is primarily concerned with bootstrapping the relationship between selectional constraints and informativity, he is most concerned with the general trends between verb classes with respect to tightness of selectional constraints. In the case of a few verbs that classically perform counter to their predicted category, Resnik falls back on arguments of aspect and manner. I will touch on these arguments in Section 5.4.5

optionality of the direct object participant role to unify the terminology within this dissertation (i.e., "object-drop" → "syntactically optional", "non-object-drop" → "syntactically obligatory"). For all these verbs, the direct object is semantically obligatory.

Syntactically Obligatory Direct Objects: *bring*, *catch*, *do*, *find*, *get*, *give*, *hang*, *have*, *hit*, *like*, *make*, *open*, *put*, *say*, *see*, *show*, *take*, *want*, *wear*

Syntactically Optional Direct Objects: *call*, *drink*, *eat*, *explain*, *hear*, *pack*, *play*, *pour*, *pull*, *push*, *read*, *sing*, *steal*, *watch*, *write*,

Lexicon 3.1: Direct Object Verbs in the Brown Corpus (Resnik, 1993)

I begin this section replicating Resnik's Experiment 1 comparing η values. The replication allows me to directly compare an analysis containing only higher frequency verbs and an analysis containing both high and low frequency verbs. Resnik compared verbs in the Brown Corpus (Francis & Kučera, 1982) while I will be comparing verbs in the BNC (see Section 2.3 for an explanation of my extraction techniques). Resnik's Brown Corpus study only uses the verbs listed in Lexicon 3.1, while I include additional verbs as shown in the expanded verb list in Lexicon 3.2. Resnik's verb choices were restricted by the smaller corpus he used in his studies in comparison to the corpus I use in my studies. There exists the possibility that the lower frequency verbs Resnik did not have corpus data for behave differently from the higher frequency verbs he did have data for. In the end, Resnik's overall results remain unchanged when tested on a larger corpus with lower frequency verbs. There is, however, a general weakening of the effect of predictability on modification rates when lower frequency verbs are included in the analysis.

Table 3.2 show the cumulative strength (the η or selectional constraint) associated with those verbs that did occur in the Brown Corpus. Figure 3.2 recasts these findings graphically to show the relative distribution of these selectional constraint values. This graphical representation make clear the average (the white dot) and the interquartile range (the

Syntactically Obligatory Direct Objects: *bring*, *catch*, *do*, *find*, *get*, *give*, *hang*, *have*, *hit*, *like*, *make*, *open*, *put*, *say*, *see*, *show*, *take*, *want*, *wear*

Syntactically Optional Direct Objects: *answer*, *approach*, *attend*, *build*, *call*, *change*, *choose*, *continue*, *cut*, *draw*, *drink*, *drive*, *eat*, *enter*, *explain*, *fly*, *follow*, *gain*, *grab*, *hear*, *judge*, *kick*, *know*, *lead*, *leave*, *lose*, *marry*, *obey*, *order*, *pack*, *paint*, *pass*, *pay*, *play*, *pour*, *print*, *pull*, *push*, *read*, *recall*, *refuse*, *remember*, *sing*, *steal*, *swallow*, *think*, *waste*, *watch*, *wave*, *write*

Lexicon 3.2: Expanded Direct Object Verb List

heavy centerline) for selectional strength. The general shape of the plot conveys the kernel density of values. That is, the ranges of values that represent a higher percentage of all the values are wider. The effect size ($d = 1.34$), in terms of standardized difference

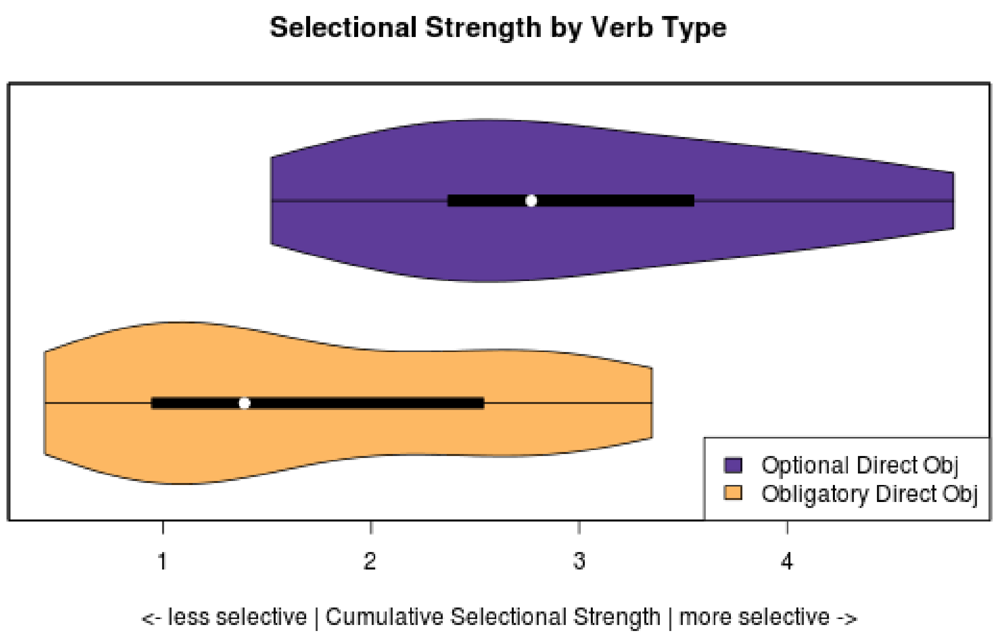


Figure 3.2: Selectional Constraint (η) in Resnik's Experiment 1 with the Brown Corpus

of means, is very large (Cohen, 1988). As should be visually apparent, syntactically optional direct object verbs have a higher average η value. Figure 3.3 shows η values for the same set of verbs in the BNC. According to the Pearson's correlation coefficient, there

Table 3.2: Verbs Used in Experiment 1 on the Brown Corpus (Resnik, 1993, pg. 138)

Object-drop verbs		Non-object-drop Verbs	
Verb	Strength	Verb	Strength
pour	4.8	*hang*	3.35
drink	4.38	*wear*	3.13
pack	4.12	*open*	2.93
sing	3.58	*say*	2.82
steal	3.52	*like*	2.59
eat	3.51	*hit*	2.49
push	2.87	*catch*	2.47
pull	2.77	*do*	1.84
write	2.54	*want*	1.52
play	2.51	*show*	1.39
explain	2.39	*bring*	1.33
read	2.35	*put*	1.24
watch	1.97	*see*	1.06
hear	1.7	*find*	0.96
call	1.52	*take*	0.93
		get	0.82
		give	0.79
		make	0.72
		have	0.43

was a positive correlation between the η values in the Brown and BNC corpora, $r = 0.68$, $p < 0.001$. The effect size for the BNC is much smaller ($d = 0.63$), although it is still a medium effect (Cohen, 1988). Comparing Cohen's d in the Brown Corpus and the BNC, we see a slight weakening of the correlation between syntactically optional direct object verbs and a high average η value (strong selectional constraint). In other words, the relationship still exists in the larger corpus but it will be diluted or not as obvious.

The strength of relationship between syntactic optionality and η is further diluted—although still present—when we add low frequency verbs to the analysis. Lexicon 3.2 shows an expanded direct object verb list. This larger verb list contains 36 additional direct object verbs that occur at least 50 times in the BNC (i.e., roughly twice as many verbs). The η values (selectional constraint) for the verbs are listed in Appendix A. Figure

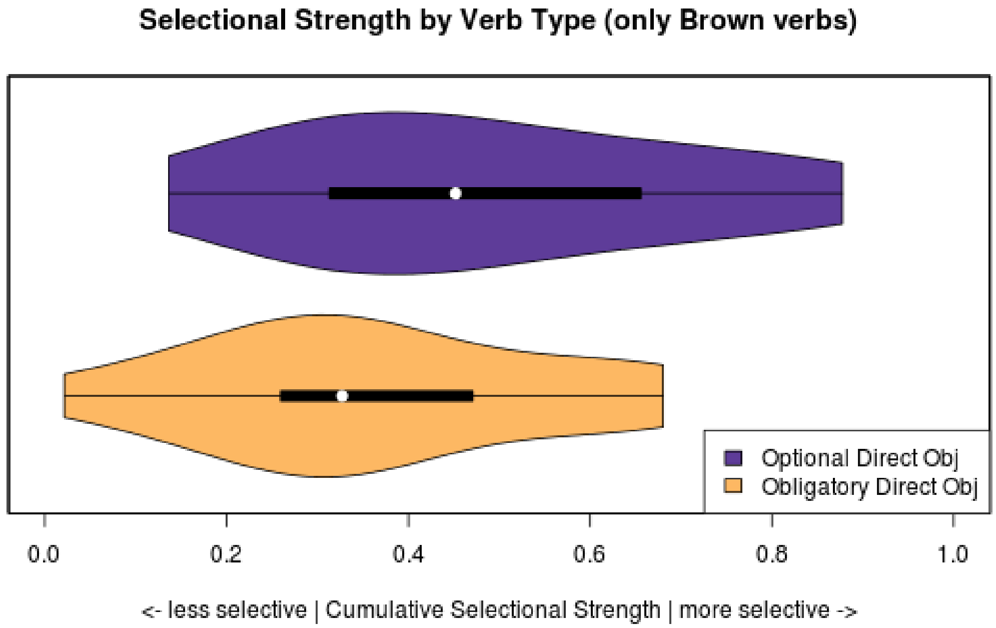


Figure 3.3: Selectional Constraint (η) Replicating Resnik's Experiment 1 with the BNC

3.4 expands the original verb list from those in Lexicon 3.1 to all those in Lexicon 3.2. The effect size using all verbs (d = 0.21) is below the rule of thumb for small effect (d = 0.25, Cohen, 1988). The shrinking effect size across the three studies (Resnik's original study, my analysis of the same verb set but using the BNC as input, and the full list of verbs in the BNC) can be partially explained in the decreased standard deviation of the obligatory verb η values. The theoretical implications of this mathematical explanation are that η value differences between obligatory and optional verbs are weaker or less determined in the larger dataset than in the smaller, high-frequency verb dataset. Even in the original dataset though, there was no single value that can be taken as a cut-off between these two verbs classes.

Resnik presents two likely possibilities for a lack of clear cut-off: (a) these corpus measures are but poor estimates of the real selectional strength values and (b) other factors are at play. While it may be true that cleaner estimates could be generated (for example,

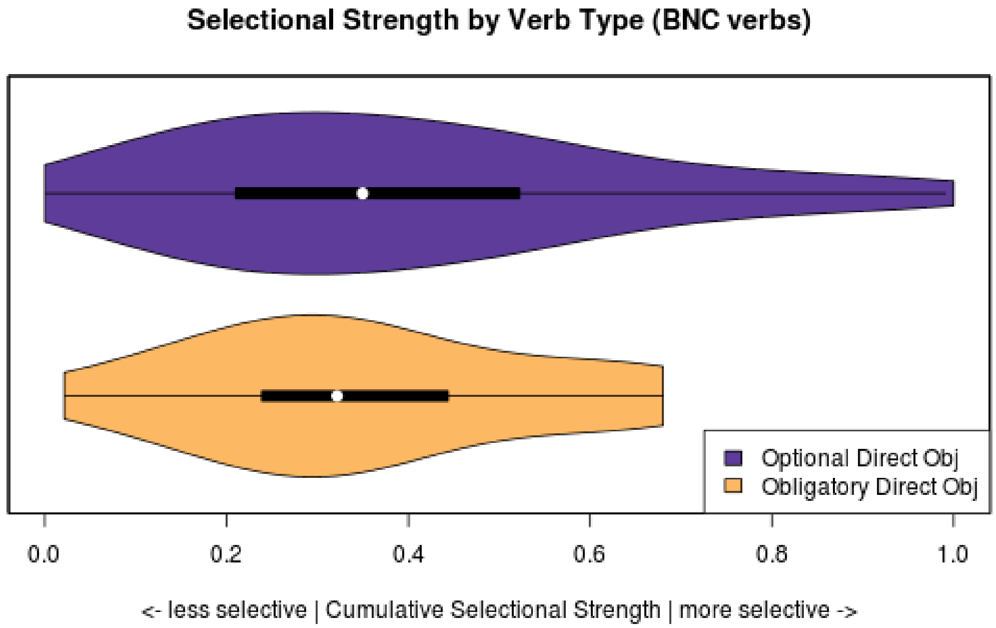


Figure 3.4: Selectional Constraint (η) Expanding on Resnik's Experiment 1 with the BNC

by using a larger corpus as I do or writing even finer grained `tgrep` patterns), I think the preponderance of the problems lies in the second possibility. The verb *devein*, while having incredibly strong selectional preference for the direct object *shrimp*, does not license direct object omission.

(103) The chef deveined shrimp.

(104) * The chef deveined.

Some factor other than aggregate selectional strength must a determining factor in direct object optionality. Resnik (1993, pg. 88) further states:

> [V]erbs do not omit their objects frequently unless they possess a high selectional preference strength. I would argue that this pattern reflects an underlying hard requirement, namely that strong selection is a necessary condition for object omission. Whatever other sources of information may be available

for inferring properties of implicit objects, selectional information carried by the verb is a prerequisite.

In short, high η value is a necessary but not sufficient condition for direct object optionality. Verbs with much higher overall selectional association (η) are more likely to allow implicit direct objects because the majority of the information conveyed by including (a highly predictable) direct object can be part of the information conveyed by the verb itself. This conclusion only addresses a general property of the lexical item. It is a context-free property (selectional promiscuity/selectional constraint) influencing a context-free property (direct object omissibility). What about measures like direct object realization (i.e., modification) that cannot be defined in the mental lexicon because they are only meaningful within a sentential context? How does selectional promiscuity of a verb or the selectional strength between a verb and some particular direct object participant role filler impact modification?

3.3 Study 1: Simple Modification Proportion

In this next study, I investigate an argument's modification rate as a function of the syntactic optionality of the argument for a verb, a verb/role filler selectional strength, and a verb's η (selectional constraint). First, I compare direct object modification rates for syntactically obligatory and optional direct object verbs. Second, I compare modification rates between verbs with high and low η values. Third, I compare modification rates for sentences with high and low verb/direct object selectional strengths. For each of these comparisons, I present a summary table of proportions and the effect size in terms of standardized mean difference between proportions. Because these tables present proportions, the cell values are dependent on each other which makes ANOVA and similar analyses invalid. The summary table provides a view of the general trend while the effect size informs us as to its general perceptibility. Within the context of these studies, we can use effect size to compare relative perceptibility between relationships. That is, the relationships with larger effect sizes should be more apparent to speakers and listeners than those relationships with medium or small effect sizes. Large effect sizes tells us that (if this is a reliable difference) interlocutors should be grossly sensitive to the difference. A small effect size means the distinction will only be obvious in larger corpora or in more controlled situations.

In terms of pragmatic processing, we should expect to see higher modification rates for more predictable direct objects. A highly predictable direct object is inherently less informative than an unpredictable direct object. As such, interlocutors would need to modify the highly predictable direct objects in order to not break the Grice's Maxim of Quantity. In contrast, van der Sluis & Krahmer (2007, 3235) found that "[t]argets ...requir[ing] more effort to refer to more often result in overspecified references". More restrictive verbs will have more and/or stronger ties to their role fillers than less restrictive verbs. As such, it should take more effort to activate the connection between less restrictive verbs and an average role filler. The increased effort will result in overspecified unpredictable direct objects.

Table 3.3 shows modification proportions by verb class. To generate the table, I calculated the modification proportion for each verb. This modification proportion is then averaged with other verbs of the same syntactic optionality class. The standardized mean difference between modification proportions has a small effect size (d = 0.25, cf. Table 3.6) according to Cohen (1988). Direct objects of obligatory verbs tend to be more modified. If we assume that syntactically obligatory verbs have weaker or less reliable connections to their direct objects than syntactically optional verbs, this table provides support for van der Sluis & Krahmer (2007)'s analysis (and the Maximum Context Hypothesis): weaker mental lexicon connections result in more often modified role fillers.

Table 3.3: Modification Proportion for Direct Objects by Verb Class (avg. by verb type)

	Optional	Obligatory
bare	0.48	0.45
modified	0.52	0.55

Table 3.4 shows modification proportions by η value. After calculating modification proportion per verb, I average across verbs using bins for high and low η value verbs. Binning η values into high and low allows me to expose a simple trend through a pivot table. I use linear mixed-effect models in a later section to analyze the measure directly. The standardized mean difference has a medium effect size ($d = 0.72$, where $d = 0.8$ is considered large). Here, we see that high η verbs (e.g., *drink* in Example 105) are more often bare than their low η counterparts (e.g., *hang* in Example 106).

Table 3.4: Modification Proportion for Direct Objects by Verb's η (avg. by verb type)

	low	high
bare	0.43	0.51
modified	0.57	0.49

(105) a. We drank *our coffee* in silence.

b. I drank *three cups of vile, greyish coffee.*

(106) a. The tailor hung *his creations* on the railings.

b. In all of them hang *net curtains.*

High η verbs are more likely to trigger event descriptions like those in Example 105a than those in Example 105b. Low η verbs are more likely to trigger event descriptions like those in Example 106b than those in Example 106a. These results again support van der Sluis & Krahmer (2007) and also have a stronger effect size than syntactically obligatory and syntactically optional verbs (cf. Table 3.6).

Table 3.5 shows modification rate by verb/direct object role filler pairwise selectional strength. As you may recall from Section 2.5.2, selectional strength is computed at the sentence token level rather than the verb type level. For the current table's analysis, I categorized the verb/filler pairs of each sentence as having a high or low selectional strength. Again, this binning helps paint a simple picture of the relationship between selectional strength and modification type. Next, I calculate the proportion of modified direct objects for low selectional strength pairs and for high selectional strength pairs.

Table 3.5: Modification Proportion for Direct Objects by Verb-Direct Object Selectional Strength (avg. by sentence token)

	low	high
bare	0.45	0.41
modified	0.55	0.59

The standardized mean difference for selectional strength is the lowest of all three tables (d = 0.11, cf. Table 3.6) and falls below Cohen (1988)'s rule of thumb for small effects. We can see from the table that event descriptions with low selectional strength between the verb and direct object role filler have a lower role filler modification rate (at 55%) than event descriptions with high selectional strength (at 59%). To make this interaction concrete, the term *coffee* has high selectional strength with the verb *drink* (as in Example 105, repeated below) and low selectional strength with the verb *order* (as in Example 107). The role filler *coffee*, in the context of *drink*, will more often be modified than in the context of *order*. As I assume selectional strength is an indicator of informativity, this table supports a Minimum Effort Hypothesis. Lower informativity (from higher selectional strength) correlates with higher modification rates.

(105′) (105a′) We drank *our coffee* in silence.

(105b′) I drank *three cups of vile, greyish coffee.*

(107) a. You erred by ordering *coffee.*

b. I ordered *two coffees with rum* at the bar...

Table 3.6: Modification Proportion for Instruments (avg. by verb type) with Cohen's *d*

	Verb Class		Verb's η		Selectional Strength	
	Optional	Obligatory	low	high	low	high
bare	0.48	0.45	0.43	0.51	0.45	0.41
modified	0.52	0.55	0.57	0.49	0.55	0.59
Cohen's *d*	0.25		0.72		0.11	

To summarize, we see that selection strength, when binned into high and low categories, is weakly (in terms of effect size) in support of the Minimum Effort Hypothesis. Not taking into account other factors, higher selectional strength between verb/direct object pairs correlates with higher modification rates. However, the other two predictors (direct object optionality and selectional constraint) are more strongly (in terms of effect size) in support of a van der Sluis & Krahmer (2007)'s analysis (the Maximum Context Hypothesis). In both cases, more predictable role fillers are more likely to be bare. In the next section, I will use linear models to allow us to analyze multiple factors at the same time to help tease apart this conflicting evidence. Even without the linear model analyses, the relative effect sizes tell us that selectional constraint (η value) is most likely the most important factor while selection strength, if actually significant, is the least important factor.

3.4 Study 2: Modification Proportion as Modulated by Informativity

General Introduction. I have shown that syntactically obligatory verbs have less predictable role fillers than syntactically optional verbs. I have also shown simple trends between the three factors of interest (all measuring event description informativity) and modification rates. What remains to be shown is which, if any, of the simple correlations discussed in the previous section significantly affect speaker choices when all factors are taken into account. How is a speaker's decision to modify or leave bare a direct object role filler impacted by their understanding of how informative or unexpected their intended event description is? According to the Maximum Context Hypothesis, speakers will tend to modify those least expected role fillers. According to the Minimum Effort Hypothesis, speakers will tend to modify those most predictable role fillers.

General Methodology. In Study 2, I combine all the predictors of interest in a single linear model to test the Maximum Context Hypothesis and the Minimum Effort Hypothesis on direct object role fillers. Linear models allows us to investigate interactions between predictors and measure finer-grained interactions than the simple pivot tables in the previous section. Specifically, a linear model could resolve the ambiguities introduced by the competing trends in Tables 3.3–3.5. More predictable event descriptions correlated with more bare direct objects in two of the analyses while more predictable event descriptions correlated with more modified direct objects in the third analysis. A linear model would allow us to include all three predictors from the three separate analyses in a single model. If the different predictors explain the same aspects of the relationship, then not all three will have significant impact on the dependent variable of modification rate. If the different predictors explain different aspects (i.e., they are mutually exclusive in some way), then multiples of them will have a significant impact on the dependent variable. Section 2.5 in Chapter 2 outlines all of the independent measures used in the linear models below.

The dependent measure is modification status (bare = 0, modified = 1). The predictors themselves have all been z-score normalized and transformed to guarantee linearity. The former will make analyzing more straight forward and the latter is a requirement of linear mixed-effect models. Following Wurm & Fisicaro (2014), predictors with high correlations were not residualized.[8] I used fully crossed fixed-effects. I specified random intercepts with respect to direct object synsets. I specified random slopes for verb diversity, direct object diversity, and verb class with respect to direct object synsets. Including other direct object synset random slopes caused the models to not converge.

I exclude random slopes for these two predictors in my models. Including random slopes and intercepts for both verb lemmas[9] and direct object synsets produced models that did not converge or were undefined due to too many degrees of freedom for the dataset. The theoretical implication for excluding these predictors is that we must assume the effect of η value and selectional strength on direct object modification to be invariant between direct object synsets.

I also assume the effect on direct object modification to be invariant between verb lemmas. In other words, for each verb lemma, the baseline modification (i.e., the intercept) for any direct object will not vary between verb lemmas. Likewise, the interaction between each predictor and a verb lemma type (i.e., the slope) will not vary over a normal distribution. As a consequence, I exclude random slopes and intercepts for verb lemmas in my models.

I pulled the same number of bare and modified sentence tokens for each verb to further control for inter-verb token variation. For instance, if a verb had 500 modified sentences and 510 bare sentences, I only used 500 of each in the final model. Pulling the same number of bare and modified event descriptions per verb reduced the likelihood that linear model would overfit the corpus. In the world of event descriptions outside of my corpus, individual verbs no doubt have strong individual biases as to how likely their direct objects are to be modified or bare. Controlling for this bias through balanced sampling means that

[8] According to previous standards, I residualized predictors with high correlations. Models using this residualized data did not differ from those reported below in any significant way.

[9] Again, I am using the term *lemma* here to mean roughly citation form, as is standard in computer science.

I do not need to include yet another independent control measure in the model.

Following the practices discussed in Barr et al. (2013), I report the most complex model that converges. Analyses were conducted using the `lme4` (Bates et al., 2013, version 0.999999-2) and `languageR` libraries (Baayen, 2011, version 1.4) for the R statistics program (R Core Team, 2012, version 2.15.1).

In Section 3.4.1, I analyze a baseline linear model and a complete linear model using the verbs listed in Lexicon 3.2. The baseline model against which my complete model will be compared includes verb diversity[10], direct object diversity, and the interaction of verb diversity x direct object diversity as predictors. Random and fixed effects for the complete model have already been discussed above. Analyzing the complete model with respect to the baseline model tells us which factors are significant and, most importantly, which factors contribute above and beyond a simple explanation prior to any novel contributions. A parallel analysis was performed using only the 34 verbs from Resnik's Brown Corpus study but only minimal differences exists between that analysis and the one presented below.

3.4.1 Study 2: A Linear Model of All Verbs

Baseline Model Results. The following models are based on a total of 67 verbs and 144,340 sentence tokens. Table 3.7 shows the baseline model, without predictors of interest. All terms are significant. The intercept and the β value for direct object diversity are positive and significant. The β value for verb diversity and the interaction of verb diversity and direct object diversity are negative and significant.

The intercept reflects the baseline bias with respect to direct object modification. The negative β value for the interaction means that verb diversity and direct object diversity

[10] Diversity is a measure similar to frequency but one that is reported to be more psychologically valid (Adelman et al., 2006). The measure reflects the number of different passages (or documents) a word occurs in for a given corpus. See Section 2.5.1 for a full explanation.

Table 3.7: Baseline Linear Model β Values for All Verbs ($n > 50$)

	Coef β	SE(β)	z	p
Intercept	0.05	0.02	2.2	**<.05**
Verb Diversity	−0.32	0.07	−4.7	**<.0001**
Direct Object Diversity	0.06	0.02	3.5	**<.001**
Verb Diversity x Direct Object Diversity	−0.24	0.06	−4.1	**<.0001**

Note: The predictors have all been z-score normalized. A positive coefficient indicates increased modification likelihood.

have an additive effect on each other. The negative β value for verb diversity's simple effect means that more diverse verbs (i.e., those used in more contexts) are more likely to precede bare direct objects. The positive β value for direct object diversity's simple effect means that more diverse direct objects (i.e., those used in more contexts) are more likely to be modified. The input data for both diversity predictors is on the same scale because both have been z-score normalized. As such, we can meaningfully compare the relative values of each to see that verb diversity is the dominant factor of these two.

Complete Model Results. Table 3.8 contains the β values for my complete model (i.e., the best model containing all of the control factors and the additional factors of interest). Of the original significant predictors, the intercept is no longer significant. The simple effect of direct object diversity and the interaction between the two diversities are no longer significant.

Verb diversity is now only significant in interactions (i.e., those between more than one predictor). Namely, we have more precisely narrowed the impact of verb diversity to certain conditions. As I mentioned above, the relative size of the β's absolute value informs us as to the relative impact of that predictor on the final modification state. Specifically, verb diversity participates in two significant interactions: with verb class and with selectional constraint. Increases in verb diversity predict lower modification rates if we

Table 3.8: Complete Linear Model β Values for All Verbs ($n > 50$)

	Coef β	SE(β)	**z**	**p**
Intercept	−0.03	0.03	−0.9	0.4
Verb Diversity	−0.07	0.16	−0.4	0.7
Direct Object Diversity	0.04	0.02	1.9	0.06
Verb Class (obligatory = 1)	0.20	0.06	3.2	**<.01**
Selectional Constraint (η)	0.06	0.01	4.3	**<.0001**
Selectional Strength	−0.16	0.05	−3.4	**<.001**
Verb Diversity x Direct Object Diversity	−0.09	0.09	−1.1	0.3
Verb Diversity x Verb Class	0.83	0.18	4.7	**<.0001**
Verb Diversity x Selectional Constraint	−0.17	0.07	−2.4	**<.05**
Verb Diversity x Selectional Strength	−0.03	0.17	−0.2	0.9
Direct Object Diversity x Verb Class	0.04	0.04	1.2	0.2
Direct Object Diversity x Selectional Constraint	−0.01	0.01	−1.2	0.2
Direct Object Diversity x Selectional Strength	−0.05	0.03	−1.7	0.1
Verb Class x Selectional Constraint	−0.03	0.03	−0.7	0.5
Verb Class x Selectional Strength	−0.24	0.09	−2.6	**<.01**
Selectional Constraint x Selectional Strength	0.07	0.01	4.9	**<.0001**

Note: The predictors have all been z-score normalized. A positive coefficient indicates increased modification likelihood.

hold either verb class or selectional constraint constant. To flip around this interaction, increases in selectional constraint predict lower modification rates if we hold verb diversity constant. The interaction between verb diversity and verb class seems to serve as a dampening effect on these other interactions for obligatory verbs. Its β value is positive, which implies modification rates go up as verb predictability increases for syntactically obligatory direct object verbs. The β values of the other significant predictors predict the opposite trend.

Verb class has a simple effect, in addition to participating in other complex effects. Obligatory verbs are more likely to accompany modified direct objects, according to the direction of the simple effect's β value. The complex effect of verb class with selectional strength push in this same direction. Namely, obligatory verbs with high selectional strength values also correlate with low modification rates. Looking across both of these verb class related effects, we see a general bias towards the Maximum Context Hypothesis. Three of the four effects are weighted such that the less predictable or weaker connected situation correlates with higher modification rates.

The simple effect of selectional strength correlates higher strength with lower modification rates. Selectional strength also participates in an interaction with selectional constraint. This interaction of selectional strength and selectional constraint has a dampening effect on most of the other significant interactions. In other predictors, increases in selectional strength or selectional constraint predict a lower probability for modification. For this dampening interaction, increases in selectional strength or selectional constraint predict a higher probability for modification. Likewise, the simple effect of selectional constraint predicts that higher values correlate with higher probability for modification.

3.4.2 General Discussion

The general trend across these verb/direct object studies is for strong mental lexicon connections (defined both in terms of selectional strength and η) to be a good predictor of low modification rates. These results cannot be fully explained in terms of raw activation of a verb and role filler. Diversity is the measure that should most directly correlate with a concept's activation level (see Adelman et al., 2006, for justification of this claim). If the underlying mechanism for increased modification was just very low activation, then we should expect to see high modification also correlate with low diversity, since low diversity is another indicator of low activation. Instead, for verbs, higher diversity tends to correlate with higher modification rates. Direct object diversity stops being a significant factor when my additional predictors are modeled. I found the simple effect of selectional strength, the simple effect of verb class, the complex effect of verb diversity x

selectional constraint, and the complex effect of verb class *x* selectional strength to predict that stronger connections lead to lower modification rates. These results align with those of van der Sluis & Krahmer (2007) that show harder to access referents tend to be overspecified. I will hold off further discussion of the implications of these patterns until after analyzing the relationship between semantically optionality, predictability, and modification rates.

The author intended for this page to be blank.

Chapter 4

Comparing Semantically Obligatory and Optional Roles

One must not put a loaded rifle on the stage if no one is thinking of firing it.

—Chekhov, letter to A. S. Lazarev (pseudonym of A. S. Gruzinsky), 1 November 1889

4.1 Overview

In this chapter, I use verb/instrument role filler pairs to investigate the Maximum Context Hypothesis and the Minimum Effort Hypothesis. The critical distinction between the two hypotheses is how speakers' modification choices are affected by their internal understanding of event description (un)expectedness. If a speaker's reaction to unexpected event descriptions is to provide more context, then the Maximum Context Hypothesis dominates. Unexpected arguments will be modified significantly more often than predictable argument. If a speaker's reaction to predictable event descriptions is to provide additional content, then the Minimum Effort Hypothesis dominates. The most predicted arguments will more often occur in modified contexts. Specifically, I will look at the relationship between semantic optionality of instruments, a verb and instrument's selectional strength with respect to each other, selectional constraint (the aggregate selectional strength for a verb), and modification.

First, the relation between semantic optionality and predictability for instrument verbs and instrument role fillers should parallel that of syntactic optionality and predictability for direct object verbs and direct object role fillers. Both of these verb class distinctions correspond to differential lexical encoding strength. Resnik (1993) argues that the mechanism driving the relationship between syntactic optionality and predictability is the (partial) encoding of predictable direct object participant roles on the verb. This encoding implies strong connections in the mental lexicon between a verb and its most predicted role fillers (or their features). Likewise, Koenig et al. (2003) argue that semantically obligatory verbs differ from semantically optional instrument verbs in terms of their semantic participant lexical encoding. Koenig et al. (2008) argue that this difference is realized (at least partially) through the additional constraints that semantically obligatory verbs impose on their instruments than semantically optional verbs. Putting these two arguments together, we will expect to find more and stronger connections in the mental lexicon between verbs and their semantically obligatory role fillers.

Second, surface realization of instrument role fillers should be impacted by the strength of connection to their verb. I argued in the previous chapter that the mechanism that ties modification of direct objects to mental lexicon connectivity is activation or accessibility. van der Sluis & Krahmer (2007) make similar arguments that decreases in accessibility are one root cause of increasing modification rates. Speakers are influenced by how difficult the direct object concept is to access or activate. Modification of instruments should be governed by the same forces. What remains to be seen is whether the strength of connection is best determined by optionality class, by aggregate predictability (i.e., η or selectional constraint), by actual predictability (i.e., selectional strength), or by some combination of the three. It is also possible that the blanket syntactic optionality of instruments or the semantic optionality for some instrument verbs fosters a different set of forces. Or it could be that these forces are only strong enough to be observed in direct object role fillers, due to their much higher predictability and usage in event descriptions.

Like with direct object role fillers, we want to compare event descriptions that represent high and low predictability environments for listeners. Ideally, I would use instrument inclusion or omission as my dependent variable. For semantically obligatory instruments, I could calculate instrument omission rates by counting the number of instrument verbs with an explicit instrument and dividing it by the total number of instrument verbs. Unfortunately, the same measure cannot be calculated for semantically optional instrument verbs because the denominator is not defined. A (semantically) optional instrument verb without an explicit instrument could either be describing an event with an implicit instrument (semantically present but syntactically absent) or be describing an event without an instrument (neither semantically nor syntactically present). The latter occurrences are outside the scope of relevant instances but cannot be distinguished from the former occurrences. Instead, I am going to gather evidence based on whether a mentioned instrument is bare or modified, following the same logic as in Chapter 3. I assume that an omitted referent is less informational[1] than an overt referent, when describing the same event. Likewise, a bare referent is less informational than a modified referent.

Predictions for the following studies align with our predictions from the previous chapter. Specifically, syntactically *obligatory*[2] direct object verbs are those verbs with *weaker* mental lexicon connection to their typical direct object. Syntactically **optional** direct object verbs are those verbs with **stronger** mental connection to their typical direct object (Resnik, 1993). In the same light, semantically **obligatory** instrument verbs

[1]See Section 1.2 for a more thorough explanation of what I mean by informational.

[2]As a caution, I warn the reader to be careful about their associations with the modifiers *obligatory* and **optional** in transitioning from Chapter 3 to this chapter.

have **stronger** mental lexicon connections to their typical instruments than semantically *optional* instrument verbs (Koenig et al., 2003, 2008; Yun, 2012). Therefore, I predict that higher η values will be associated with semantically obligatory verbs just as higher η values are associated with syntactically optional verbs. Both of these verb classes have stronger mental lexicon connections to their respective typical participant roles.

Evidence from the previous chapter does not directly inform predictions for instrument modification rates. I remind the reader that instruments of semantically optional instrument verbs are categorically different from instruments of semantically obligatory instrument verbs and direct objects of either verb class because the former is the only semantically optional participant role. In the other three cases, the participant role – whether syntactically present or not– is still semantically obligatory. In the case that instruments pattern like direct objects, stronger mental lexicon connections will correlate with more likely bare instruments. The underlying mechanism is strength of connection and activation, like with direct objects. In the case that instruments pattern opposite direct objects, connections between semantically optional instrument verbs and their instruments are sufficiently weak that there is no differential accessibility between more frequent and less frequent pairs. Namely, the cost of activating the connection between a semantically optional verb and instrument is hard enough that it swamps any difference in costs between more frequent and less frequent instruments. Direct object role fillers are much more frequent in normal speech than instrument role fillers. In this second possible outcome, the baseline connections for direct objects are already strong and easy to activate. In contrast, activating any instrument role filler would be more difficult and that activation effort would preclude finer distinctions of activation strength from being relevant (like those distinctions between semantically obligatory and semantically optional roles).

Before any further studies, I provide more concrete examples across a range of η values (selectional constraint) and selectional strengths in Section 4.1.1. These examples are intended to develop the reader's intuitions as to what 'high' and 'low' values relate to in the later studies. I correlate selectional preference and semantic optionality in Section 4.2. Section 4.3 introduces the first study to correlate instrument modification rate with strength of connection in the mental lexicon. Section 4.4 uses the more sensitive approach of linear modeling to correlate modification rate with verb class, η values, and selectional strength.

4.1.1 Instrument Selectional Constraint and Selectional Strength

As in Section 3.1.1, I present four representative verbs of high and low η values with several instruments of high and low selectional strength to provide a better understanding of the range of selectional strength for a verb and how they interact with selectional constraint. These verbs and their related instruments are in Table 4.1. The η value for each verb ($0 <= \eta <= 1$, $M = 0.38$, $SD = 0.19$) is in parentheses. The verbs *present* and *threaten* are on the lower end of the range of η values. The verbs *prod* and *lift* are on the higher end. Of the four verbs, *prod* is the only semantically obligatory instrument verb.

Table 4.1: High and Low η Verbs with High and Low Selectional Strength Instruments

Low Selectional Str.	Verb (η)	High Selectional Str.
cup book	*present* (0.11)	cheque
bankruptcy claw	*threaten* (0.2)	knife
weapon	*prod* (0.72)	stick
smile	*lift* (0.8)	hand

Verbs like *threaten* and *present* conceptually allow for a wide range of instruments but do not require an instrument. For *threaten*, *knife*-like instruments[3] have a particularly high selectional strength. This measure (as defined in Section 2.5.2) quantifies how strongly a particular instrument type is associated with a particular verb in contrast with all other verbs. These *knife*-like instruments are both very frequent with *threaten* and relatively infrequent as instruments of other verbs. In contrast, low selectional strength instruments (e.g., *claw* and *bankruptcy*[4]) are either infrequent as instruments of *threaten*, frequent as instruments of other verbs, or both.

At the other end of the η scale, verbs like *prod* and *lift* have a much narrower range of instruments. Intuitively, the typical instruments of a verb like *prod* fall into a smaller set than the instruments of a verb like *threaten.* In Section 4.2, I will explicitly test this intuition that instruments of semantically obligatory verbs have stronger selectional constraints (i.e., higher η values). For now, it is more important to note that, while the instrument of *prod* is semantically obligatory, the instrument of *lift* is semantically optional. Strong selectional constraints cannot be the sole determiner of semantic optionality.[5] If this were the case, we should be able to sort all verbs by their η value and set a threshold that exactly separates the semantically obligatory verbs from the semantically optional verbs. The lessened diversity of instruments for high η verbs can also be observed when comparing the high and low selectional strength instruments for *prod*. An instrument

[3] As covered in Chapter 2, I do not use the instruments themselves in these measures. Instead, I use the strongest WordNet synset associated with a noun, abstracting away from the word to the concept.

[4] I do not distinguish between the senses of $threaten_{claw}$ and $threaten_{bankruptcy}$. As covered in Chapter 2, I am using the verb lemma rather than the verb synset as the coin of the realm for practical and theoretical reasons. On the theoretical side, the results would be very sensitive to the exact definition used to determine when a verb sense was the same and, thus, which selectional strengths should be merged together to create a composite η score. Evaluating the different approaches to lumping or splitting verb senses constitutes a dissertation in its own right. On the practical side, the dataset used for calculating selectional strength and η would need to be significantly larger than what I have used. Muddying the distinct ηs and selectional strength measures of different verb senses through using a single verb lemma should only increase the chance of false negative results rather than increase the chance of false positive results.

[5] Likewise, η values alone cannot be used to distinguish syntactically obligatory and optional verbs.

with one of the lowest selectional strengths for *prod* (*weapon*) is conceptually still very similar to an instrument with one of the highest selectional strength (*stick*). In other words, either *weapon*s serve as a natural super category for *stick*s or *sticks* share many properties with typical *weapons*. In contrast, the same cannot be said for the high and low selectional strength examples for the low η value verbs.

Figure 4.1 makes the intuitions described above more concrete. In the top graphs, you can see the much longer tail of possible synsets to co-occur with the verbs with lower η values. This is evident in the larger number of bars across the x-axis that are associated with a probability. In more technical terms, each bar across the x-axis is a WordNet synset used as instrument role fillers for each verb. The y-axis shows the proportion of occurrences for a particular verb of a particular instrument synset. Hence, the height of the bar conveys how often a synset occurs. The synsets have been sorted by frequency from left (highest) to right (lowest) per verb. Looking at the leftmost bars, we can determine the proportion of event descriptions for the given verb that evoke the most frequent synset.

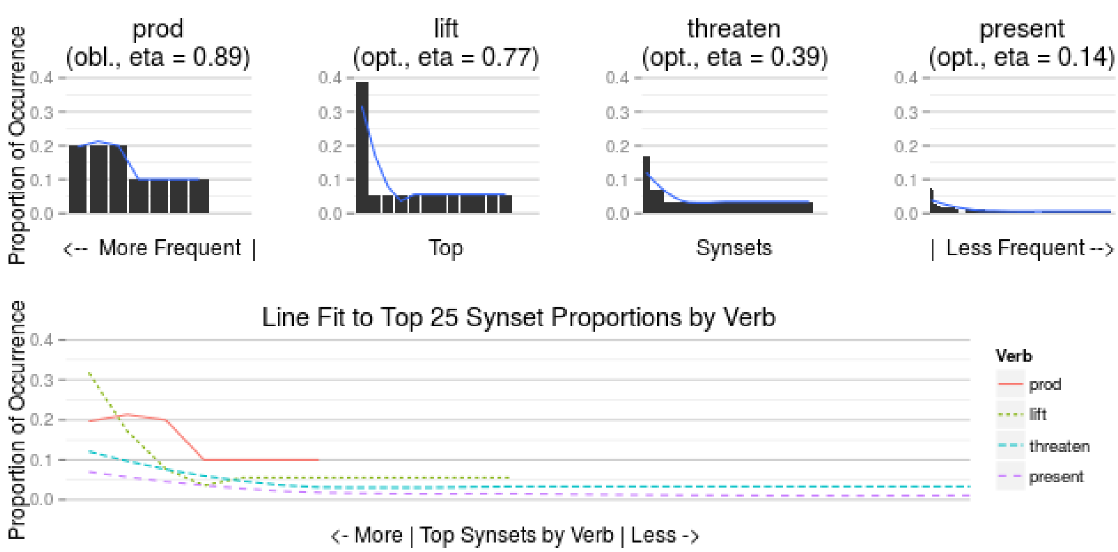


Figure 4.1: Proportion of Occurrence for Instrument Synsets with Sample Verb

The bottom graph combines the smoothed trend line for each of the four verbs to facilitate direct comparison of mention proportions. In this graph, it is clear that synset proportions for *prod* and *lift* are higher than for either *threaten* or *present*. In more technical terms, the majority of the probability mass for the high η verbs is concentrated on the left side of the graph (and associated with a smaller number of distinct synsets) while the probability mass for the low η verbs is more evenly spread across the graph (and associated with a larger number of distinct synsets).

Once again (compare with Figure 3.1 in Section 3.1.1), the majority of the probability mass for the high η verbs is concentrated on the left side of the graph (and associated with a smaller number of distinct synsets) while the probability mass for the low η verbs is more evenly spread across the graph (and associated with a larger number of distinct synsets). The high η values of *lift* and *prod* are achieved through a small number of very frequent concepts. The verb *lift* has one clear dominant synset which accounts for 50% of all its instrument mentions. The top three synsets for *prod* account for roughly 60% of all its instrument mentions. In comparison, *threaten* and *present*[6] are associated infrequently with a larger number of concepts. The top three instruments for either of the low η verb is only going to account for 30% (*threaten*) to 17% (*present*) of the total occurrences.

4.2 Selectional Preference and Semantic Optionality

My first instrument study investigates selectional preference and semantic optionality. The first goal of this study is to confirm corpus distribution parallels between direct object and instrument mentions. Parallel corpus distributions would mean that our intuitions about selectional constraint, selectional strength, and predictability should map between direct object and instrument event descriptions. Resnik (1993) provides evidence that

[6] A quick analysis of the four verbs presented may lead the reader to conclude that higher η verbs describe more concrete, physical actions while lower η verbs describe more abstract actions. As quick counter-examples, the verbs *hang* and *mop* have low η values while the verbs *indicate* and *align* have high η values. Appendix A provides a complete listing of all verbs and their η values.

there is a correlation for direct objects and their verbs between selectional preference and syntactic optionality. Much like my own studies in Chapter 3 that replicate the results on a larger corpus, Resnik uses the notion of relative entropy (see Section 1.2.1) to quantify the predictability of a given direct object given a verb. This quantification, when generalized across all direct object participant role fillers for a given verb, provides a useful metric for describing how reliably an interlocutor could correctly guess a direct object given a verb and how specific those direct objects tend to be to that verb in particular. For instance, the verb *enter* (η = 0.12 on a scale of 0–1) has relatively unpredictable role fillers from a wide range of concepts. The verb *drink* (η = 0.88 on a scale of 0–1) has an intuitive small class of role fillers that commonly serve as the direct object.

Verbs with much more predictable direct object fillers are much more likely to allow direct object omission. One of the possible arguments for why this correlation between predictability and omissibility occurs (and is even able to occur without increased message failures) is that verbs with sufficiently predictable direct objects lexically encode properties of those direct objects. Even when a direct object is omitted for these verbs, interlocutors still have strong, experience-based biases for what those fillers are.

Koenig et al. (2003) used a filler-gap study to show that participants have clear expectations for occurrence of instrument participant roles with semantically obligatory verbs but not necessarily with semantically optional verbs. These expectations are rooted in lexical encoding for the verb of these instrument types. Yun et al. (2006) found faster reading times for instruments (e.g., *spear*) when following semantically obligatory instrument verbs (e.g., *jab*) than when following semantically optional instrument verbs (e.g., *attack*). Finally, Bienvenue et al. (2007) found more anticipatory looks to instruments following semantically obligatory instruments than semantically optional instruments using a visual world paradigm.

All three of these studies underscore the stronger lexical encoding of instruments for the semantically obligatory verbs. If we assume a parallel underlying mechanism between direct object participant role fillers and instrument participant role fillers (which I have realized in terms of strength of connection in the mental lexicon), semantically obligatory instrument verbs should have much higher selectional strengths and stronger selectional constraints than their semantically optional counterparts. To test this hypothesis, I have compared η values by both token and type[7] for the verbs listed in Lexicon 4.1. These

Obligatory Instruments: *beat*, *bombard*, *build*, *cover*, *drive*, *hit*, *illustrate*, *mark*, *mop*, *pick*, *prod*, *push*, *strike*, *stroke*, *sweep*, *wipe*, *fill*

Optional Instruments: *acquire*, *add*, *align*, *answer*, *approach*, *arrange*, *attack*, *become*, *begin*, *break*, *bring*, *buy*, *carry*, *catch*, *celebrate*, *charge*, *close*, *combine*, *compare*, *complete*, *conclude*, *confirm*, *connect*, *console*, *contain*, *create*, *describe*, *drink*, *drop*, *end*, *entertain*, *enter*, *establish*, *examine*, *exchange*, *eye*, *find*, *finish*, *fix*, *follow*, *form*, *get*, *handle*, *help*, *hold*, *identify*, *include*, *increase*, *indicate*, *join*, *keep*, *kick*, *kill*, *launch*, *lead*, *leave*, *lift*, *link*, *look*, *lose*, *make*, *meet*, *mix*, *open*, *pass*, *play*, *present*, *press*, *produce*, *prove*, *put*, *raise*, *reach*, *read*, *receive*, *reflect*, *replace*, *report*, *represent*, *return*, *reward*, *rub*, *score*, *see*, *send*, *set*, *shake*, *show*, *speak*, *spray*, *stand*, *start*, *study*, *supply*, *support*, *take*, *tell*, *threaten*, *touch*, *treat*, *turn*, *view*, *watch*

Lexicon 4.1: BNC Instrument Verbs With Frequencies Greater than 10

verbs were chosen as the subset of those verbs listed by Koenig et al. (2008) that occurred in the BNC with a frequency greater than 10. This same set of verbs will be used for all the instrument verb studies that follow in this chapter. When similar comparisons as I run below were calculated using slightly lower and higher frequency cut-offs, the general patterns remained the same.

First, I compared η values between semantically obligatory and semantically optional instrument verbs by token count. Figure 4.2 depicts the distributional differences between the verb classes. The x-axis plots η as normalized from zero to one. An unpaired Wilcoxon

[7] Analyzing by verb type helps us discover if there are certain idiosyncrasies specific to some verbs but not others. Analyzing by event description tokens helps us see the general trends in terms of frequency or exposure.

rank sum test comparing ηs between verb classes showed there was a significant effect of semantic optionality, $p < 0.001$. The mean for semantically optional instrument verbs ($M = 0.56$) was lower than for semantically obligatory instrument verbs ($M = 0.60$). A Wilcoxon test was required because the data did not meet the independence requirement of a t-test as values were drawn from sentence tokens with verbs repeated across sentences. The number of data points was large enough that I did not need to apply continuity correction to smooth out the data. The effect size ($d = 0.77$) measures the standardized difference between means. The mean η value of semantically obligatory verbs is 0.77 standard deviation units higher than that of semantically optional verbs. Using Cohen (1988)'s rule of thumb, this difference is just shy of a large effect ($d = 0.8$). These results parallel with my results in Chapter 3. Theory-based reports of syntactically optional roles and semantically obligatory roles claim stronger stronger mental lexicon connections between verb and participant role than their respective contrastive verb class. Corpus-based reports for syntactically optional verbs—and now semantically obligatory verbs—show more predictable verb/participant role pairings.

I also compared η values by verb type to remove any frequency bias inherent in the verb token analysis. In a verb token analysis, the selectional strength of higher frequency verbs would be sampled more often than low frequency verbs. If an effect were only present in high frequency verbs, the effect could erroneously be assumed to be verb-general. The effect size ($d = 0.61$) is slightly lower than by verb token but still above what Cohen (1988) considers a medium effect ($d = 0.5$). Figure 4.3 depicts the distributional differences for verb types. A Welch two sample t-test showed a significant difference in the mean η values between verb classes, $t(21) = 2.26, p = 0.03$. The mean for semantically optional instrument verbs ($M = 0.60$) was lower than for semantically obligatory instrument verbs ($M = 0.62$), just as when comparing by tokens. I used a Welch two sample t-test because, while the dataset satisfies the independence assumptions of a t-test, I could

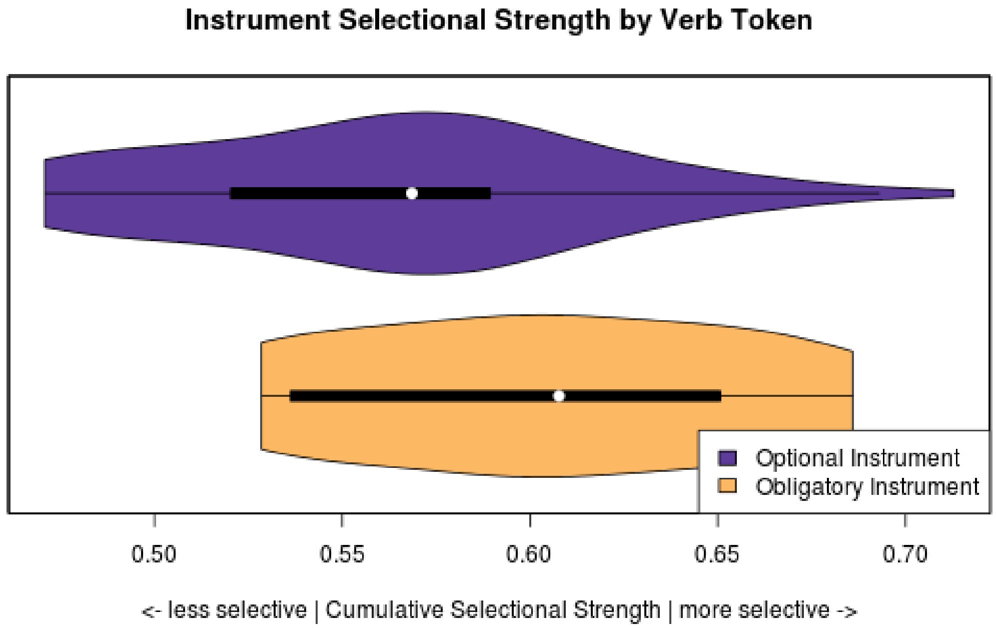


Figure 4.2: Cumulative Selectional Strength for Instrument Verb Tokens with the BNC

not assume that the variances of η values for semantically obligatory and semantically optional instrument verbs would be the same. First, each verb's η, which is only sampled once, should be independent of other η values. Further, a Shapiro-Wilk test supports that the data was sampled from a normal distribution, $p = 0.65$, making it safe to perform a *t*-test.

To summarize, I have extended the corpus results from Resnik (1993) and corroborated –also using corpus results– the psycholinguistic evidence from Yun (2012) that support correlation between mental lexicon connections and selectional preference. Both of these studies show stronger or more defined mental lexicon connections align with stricter selectional preferences in a verb. These results do not specifically speak to modification rates and informativity but they do help shore up our previous assumptions about the relationship between predictability, selectional strength, and lexical encoding of participant role information.

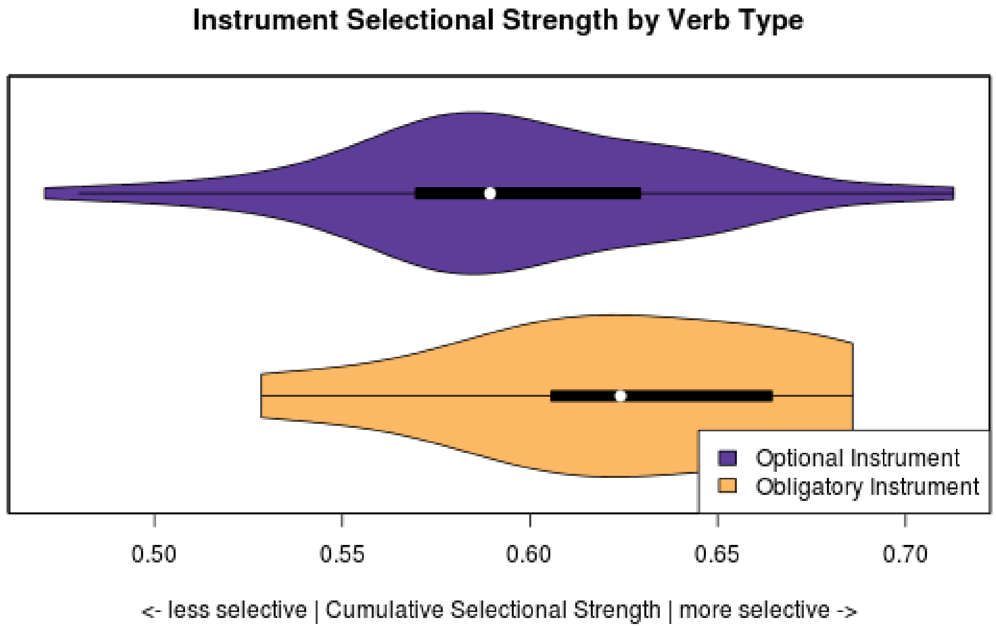


Figure 4.3: Cumulative Selectional Strength for Instrument Verb Types with the BNC

4.3 Study 3: Simple Modification Proportion

In this next study, I investigate the relationship between modification proportion and semantic optionality, modification proportion and selectional strength, and modification proportion and selectional constraint (η). First, I compare instrument modification rates for semantically obligatory and semantically optional instrument verbs. Brown (1985) and Brown & Dell (1987) present evidence that atypical instruments are more often mentioned than typical instruments in a story re-telling task. They argued that interlocutors, when asked to re-tell a story that they had been presented, preferentially mentioned atypical instruments because it was more informative. This view can be explained by Grice's Maxim of Quantity. A speaker should be verbose to the extent that pertinent information is always present. As atypical instruments are very informative, they would preferentially be included when other factors favor their exclusion.

However, with respect to instrument modification, one would expect instrument fillers to be modified so as to maximize informativity. Mentioning an instrument after a semantically obligatory verb is less informative than mentioning an instrument after a semantically optional verb. Semantically obligatory verbs always requires an instrument be part of the described event whereas semantically obligatory verbs do not entail the presence of an instrument. Mentioning an instrument for this latter class of verbs changes the underlying event structure to the extent that there is a new participant. As such, mentioning an instrument after a semantically optional verb is inherently informational whereas one must modify an instrument after a semantically obligatory verb to be similarly informative. This line of argument predicts that semantically obligatory verbs should be more likely to have a modified instrument role filler.

Pulling in the opposite direction, Koolen et al. (2011, pg. 3235) found that participants tend to overspecify those references that "require more effort to refer to". Because semantically optional verbs have categorically weaker mental lexicon connections to their instruments, one would expect them to "require more effort to refer to", resulting in higher overspecification rates. Table 4.2, which shows modification proportions by verb class, provides evidence in support of the latter argument. The standardized difference between mean modification proportions (effect size) is medium ($d = 0.70$, Cohen, 1988). The strength of connection in the mental lexicon between a verb and instrument filler seems to be an important factor in a speaker's decision to modify the instrument.

Table 4.2: Modification Proportion for Instruments by Verb Class (avg. by verb type)

	Optional	Obligatory
bare	0.51	0.60
modified	0.49	0.40

The previous studies in this chapter have all focused on the categorical distinction between verb classes. Perhaps a quantitative measure that correlates with verb class is a better indicator than the categorical verb classes. Namely, it may not be verb classes per se that are predictors of modification. It may actually be some underlying quantitative measure that strongly correlates with verb class. Semantically obligatory verbs, as a class, have stronger selectional constraints (higher η values) than semantically optional verbs (see Section 4.2). Semantically obligatory verbs, due to their higher η values, also tend to have higher selectional strength with their instruments (see Section 1.2.1 in Chapter 2 for a discussion of how selectional strength determines η value). The categorical distinction of verb class could be serving as a proxy for a different underlying mechanism: a verb's selectional constraint (η value) or the probabilistically quantified strength of connection between a verb and role filler (selectional strength). Thus, the effects on modification proportion that I have, so far, attributed to semantic optionality could in fact be a general, quantitative property of the η value rather than a categorical property of verb class. The effects could also be attributed to more specific quantitative properties of a particular verb and instrument pair (e.g., their selectional strength). In other words, there are two important properties of semantically obligatory instrument verbs: a high η value and a smaller, more semantically coherent set of typical instruments.

If the categorical difference in modification rates (between semantically obligatory and semantically optional verbs) is really a reflex of high (or low) η values (selectional constraint)[8], then we should see the same modification patterns when we bin event descriptions by the η values of the verb. Table 4.2 shows modification status by high/low η value. If the categorical difference is a reflex of instrument predictability, we should see the same modification patterns when we bin event descriptions by selectional strength. Table 4.4 shows modification status by high/low selectional strength. The effect sizes

[8]This reflex can be mechanistically explained as a smaller activation energy requirement in the mental lexicon for verbs with higher η values. A smaller activation energy requirement means that these concepts are inherently easier to access and/or activate.

Table 4.3: Modification Proportion for Instruments by Verb's η (avg. by verb type)

	low	high
bare	0.53	0.51
modified	0.47	0.49

Table 4.4: Modification Proportion for Instruments by Verb-Instrument Selectional Strength (avg. by verb type)

	low	high
bare	0.51	0.53
modified	0.49	0.47

(standardized difference of means) for low vs. high η verbs (d = 0.17) and for low vs. high verb-instrument selectional strength pairs (d = 0.11) are both below what Cohen (1988) considers a small effect (d = 0.25).

Table 4.5: Modification Proportion for Instruments (avg. by verb type) with Cohen's d

	Verb Class		Verb's η		Selectional Strength	
	Optional	Obligatory	low	high	low	high
bare	0.51	0.60	0.53	0.51	0.51	0.53
modified	0.49	0.40	0.47	0.49	0.49	0.47
Cohen's d	0.70		0.17		0.11	

I have aggregated the three comparisons from Tables 4.2–4.4 in Table 4.5 with the addition of effect size to bring to light two interesting results. First, the strength of connection between a verb and its instrument filler (selectional strength) seems a more important correlate to modification than the average expected predictability that there will be an instrument role given the verb (selectional constraint/η value). Namely, modification trends for event descriptions with semantically obligatory and semantically optional verbs parallel the trends for event descriptions with high and low selectional strength. It appears that the selectional strength component for semantically obligatory verbs is a driving factor in the former modification trends and not the verb's average predictability of its instruments (as would be supported by a parallel with the modification trends by η value). Second, the difference in modification rates between role fillers with semantically optional verbs (49% modified/51% bare) and semantically obligatory verbs (40% modified/60% bare) is much larger than the difference for role fillers with low selectional strength (49% modified/51% bare) and high selectional strength (47% modified/53% bare). This disparity is reflected both in the absolute differences (i.e., 60% ↔ 51% bare vs. 53% ↔ 51% bare) and in the effect sizes (d = 0.70 vs. d = 0.17, see Table 4.5). Assuming this higher disparity is a true difference between factors, one would expect to see it reflected in linear model coefficients in the next section. The larger disparity should correlate with a larger β value for semantic class than for selectional strength or correlate with the β value for semantic class being significant while the β value for selectional strength not being significant. In the same vein, one would expect selectional constraint (η) to either be non-significant or have the opposite sign on its coefficient, reflecting the opposite trend seen in Table 4.3.

4.4 Study 4: Modification Proportion as Modulated by Informativity

In Study 4, I look at modification status as modulated by informativity using linear models. Linear models allows us to investigate interactions between predictors and measure finer-grained interactions than the t-tests and Welch tests performed in the previous section. Specifically, a linear model could resolve the ambiguities introduced by the competing trends in Tables 4.2–4.4. Section 2.5 in Chapter 2 outlines all of the independent measures used in these models. The dependent measure is modification status (bare = 0, modified = 1). The predictors themselves have all been z-score normalized and transformed to guarantee linearity. The former will make analyzing more straight forward and the latter is a requirement of linear mixed-effect models. Following Wurm & Fisicaro (2014), predictors with high correlations were not residualized.[9] I used fully crossed fixed-effects. I used fully-specified random slopes and random intercepts for instrument synsets, when possible. Including random slopes and intercepts for both verb lemmas and instrument synsets produced models that did not converge or were undefined due to too many degrees of freedom for the dataset. By excluding random slopes and intercepts for verb lemmas, I assume the effect on instrument modification to be invariant between verb lemmas. In other words, the baseline modification rate (i.e., the intercept) for any instrument will not vary between verb lemmas. A particular instrument will be modified on average at the same rate regardless of any idiosyncrasies of the verb, if we control for all other factors. Likewise, the interaction between each predictor and a verb lemma type (i.e., the slope) will not vary over a normal distribution. There is nothing idiosyncratic about a verb lemma itself that will change the rate at which its instrument role fillers are modified. Any changes in modification probability will purely be a function of the factors of interest and control factors.

Following the practices discussed in Barr et al. (2013), I report the most complex model that converges. Patterns for simpler models generated according to the practices discussed by Baayen (2008) were also similar to those reported below. Analyses were conducted using the `lme4` (Bates et al., 2013, version 0.999999-2) and `languageR` libraries (Baayen, 2011, version 1.4) for the R statistics program (R Core Team, 2012, version 2.15.1).

The baseline model against which my model will be compared includes verb diversity[10], instrument diversity, and the interaction of verb diversity x instrument diversity as

[9]According to previous standards, I residualized predictors with high correlations. Models using this residualized data did not differ from those reported below in any significant way.

[10]Diversity is a measure similar to frequency but one that is reported to be more psychologically valid (Adelman et al., 2006). The measure reflects the number of different passages (or documents) a word occurs in for a given corpus. See Section 2.5.1 for a full explanation.

predictors. Table 4.6 shows the significant β values for this models. The intercept reflects

Table 4.6: Baseline Linear Model βs for Significant Predictors

	Coef β	SE(β)	z	p
Intercept	−0.09	0.04	−2.0	**<.05**
Verb Diversity	−0.02	0.04	−0.5	0.6
Instrument Diversity	−0.07	0.03	−2.4	**<.05**
Verb Diversity x Instrument Diversity	0.03	0.03	1.0	0.3

Note: The predictors have all been z-score normalized. A positive coefficient indicates increased modification likelihood.

the baseline bias with respect to instrument modification. The negative β value means that there is a natural bias for instruments to be bare. The base rate of instrument role filler modification in the corpus (by both sentence token and verb type) is 48%. Instrument diversity is also significant. Its negative β value means that more diversely used instruments are more likely to be bare. More context-specific or lower use instruments are more likely to be modified.

By comparison, Table 4.7 shows the significant βs for my model of interest. The negative β value for the intercept again means that there is a bias for instruments to be bare. New predictors that turned out to be significant include verb class (i.e., semantic optionality), the interaction of instrument diversity and selectional strength (between the verb and instrument), and the interaction of verb class and selectional strength.[11]

The verb class categories were coded such that semantically optional verbs were zero and semantically obligatory verbs were one. The significant, negative β value therefore implies that instruments following semantically obligatory verbs are more likely to be bare. Returning to examples from Table 4.1, instruments following *prod* are more likely

[11]Residualizing the predictors results in these two latter interactions showing up as a simple effect of selectional strength. This non-residualized model is slightly more refined in that we now better understand how selectional strength relates to modification rate.

Table 4.7: Complete Linear Model β Values

	Coef β	SE(β)	z	p
Intercept	−0.28	0.09	−3.2	**<.01**
Verb Diversity	−0.04	0.08	−0.5	0.6
Instrument Diversity	−0.03	0.05	−0.7	0.5
Verb Class (obligatory = 1)	−0.20	0.09	−2.2	**<.05**
Selectional Constraint (η)	0.07	0.08	0.9	0.4
Selectional Strength	−0.07	0.07	−1.0	0.3
Verb Diversity x Instrument Diversity	0.02	0.03	0.7	0.5
Verb Diversity x Verb Class	−0.03	0.08	−0.5	0.6
Verb Diversity x Selectional Constraint	−0.07	0.05	−1.5	0.1
Verb Diversity x Selectional Strength	0.04	0.05	0.8	0.4
Instrument Diversity x Verb Class	0.04	0.05	0.8	0.4
Instrument Diversity x Selectional Constraint	−0.05	0.04	−1.3	0.2
Instrument Diversity x Selectional Strength	0.10	0.03	3.1	**<.01**
Verb Class x Selectional Constraint	0.03	0.08	0.4	0.7
Verb Class x Selectional Strength	−0.20	0.07	−2.8	**<.01**
Selectional Constraint x Selectional Strength	−0.05	0.05	−1.0	0.3

Note: The predictors have all been z-score normalized. A positive coefficient indicates increased modification likelihood.

to be bare than instruments following *lift*. These results repeat the pattern depicted in Table 4.2 of Study 3 in this chapter.

The two significant interactions with selectional strength balance each other. The stronger of the two interactions (i.e., with verb class) aligns with the general trends seen throughout. Less connectivity in the mental lexicon or less accessibility correlates with higher modification. The weaker interaction (i.e., with instrument diversity) dampens this effect. Namely, very diverse instruments are less impacted by this general trend.

First, I will unpack the interaction between selectional strength and verb class. For semantically obligatory verbs, the higher the selectional strength between a verb and instrument, the more likely the instrument will be bare. This link between selectional strength and modification rate is not present for semantically optional verbs, as semantically optional verbs are coded as zero. The zero coding, when multiplied by the coefficient, always returns a zero for semantically optional verbs. This interaction can also be interpreted from the perspective of the instrument. For instruments of similar selectional strength co-occurring with a semantically obligatory verb and a semantically optional verb, respectively, the instrument following a semantically obligatory verb is more likely to be bare.

Next, I will unpack the interaction between selectional strength and instrument diversity. For a given instrument diversity, higher selectional strength leads to higher modification rates. Contrawise, for a given selectional strength, higher instrument diversity leads to higher modification rates.

4.4.1 General Discussion

To summarize, the general trend for instruments is the same as the general trend for direct objects. Stronger connections between a verb and its participant role filler are correlated with low modification rates. The types of verbs that generally have stronger lexical encoding for their instrument participant role fillers (semantically obligatory verbs) are generally more likely to be followed by bare instruments. When a particular verb/instrument pair has a strong selectional strength, the instrument is more likely to be bare than if the verb/instrument pair has a weaker selectional strength. In order for theories of argument representation to be able to accommodate these two effects, they must be able to encode the general and specific predictability of verb/filler argument pairs. The general effect of a verb's class and the specific effect of selectional strength are additive in some way. In a strictly categorical explanation of my results, only verb class predictors would be significant. In a strictly quantitative explanation, only scalar predictors would be significant. As both types of predictors are significant, my results appear to be caused by a combination of categorical effects and quantitative activation effects.

These results do not distinguish between a listener-oriented explanation and a speaker-oriented explanation for modification. Under a listener-oriented explanation, modifying an instrument serves to explain the connection between the verb and instrument. The speaker is trying to provide more context for the listener to help him process or integrate the event description. Under a speaker-oriented explanation, modifying an instrument is a result of mechanistic retrieval difficulties. I will use the act of describing a coffee stirring event to help illustrate the underlying differences between these explanations. Under a listener-oriented explanation, stirring coffee with a spoon is a fairly self-evident act. Most interlocutors should be comfortable with the mechanisms and motions involved in such an activity. However, if one were to stir coffee with a chain-saw, it could be useful to explicate how the act occurred. Was the chain-saw being used for its rod-like properties, were the running blades used to churn the drink, or was the engine used to vibrate the drink into a solution? Under a speaker-oriented explanation, retrieving a typical instrument like a spoon will be fast and easy. Retrieving an atypical instrument like a chain-saw will be hard. In order to facilitate retrieval of the word chain-saw, a speaker may activate related properties of the instrument, which, due to their heightened activation, are also more likely to get mentioned.

Selectional strength surfaces as a better predictor than a verb's selectional constraint (η) of modification. A verb's selectional constraint is not part of any significant predictor for instrument modification and is only significant in complex interactions for direct object modification. We should not be surprised by this general bias for results from selectional strength but not selectional constraint. A verb's selectional constraint essentially summarizes predictability across tokens while selectional strength is about a particular pair of tokens.

Finally, there are differences between the significant predictors for direct object modification and instrument modification. Some of these differences may reflect differences between syntactic and semantic optionality. Investigating other roles that differ with respect to their syntactic and/or semantic optionality should elucidate these differences. Analyzing these sames roles in other languages will also help clarify how many of these differences are inherently tied to the role, are underlyingly specific to English, or are an accident of the particular corpus analyzed. Other differences may be caused by the relatively larger and more diverse sample of direct object role fillers and event descriptions than instrument role fillers and event descriptions. Again, these differences could be inherent to some underlying differences between instruments and direct objects or an effect of the corpus. In the next chapter, I discuss methods for fixing errors that could have crept into the analysis, methods for expanding the semantic model used to represent the verbal and role filler concepts, and other approaches that should generally clarify some of these underlying differences.

The author intended for this page to be blank.

Chapter 5

Implications and Future Research

How pregnant sometimes his replies are!

— Polonius, *Hamlet*, Act II, Scene ii

My results have implications for models of language production and language comprehension. In Section 5.1, I summarize my findings. I then frame the findings within the larger domain of language production in Section 5.2. With respect to language comprehension, I address computational models of semantic relatedness and psycholinguistic models of processing in Section 5.3. I close in Section 5.4 with an overview of directions for future work.

5.1 Overview of Findings

Overall, I found lower predictability for an event description correlated with higher modification rates of the role filler. My primary findings are from the statistical models of role filler modification in Chapters 3 and 4. In both chapter, my predictors of interest are three measures of predictability from the speaker's perspective: role filler optionality, verb/role filler selectional strength, and the verb's overall selectional constraint (η). The explanatory purpose of these models is to tie predictability to role filler modification.

In Chapters 3 and 4, I tied the higher predictability of direct object role fillers and instrument role fillers, respectively, to lower modification rates. The first measure of predictability (role filler optionality) was the *syntactic* optionality of the direct object given the verb or the *semantic* optionality of the instrument given the verb, depending on the chapter. The second and third measures of predictability are identical between the chapters. The verb/role filler selectional strength measures relative predictability of a particular role filler given the verb. A verb's selectional constraint measures the aggregate selectional strength for the verb across all role fillers and is intended to quantify how predictable in general its role fillers are.

I used linear mixed effect models with these and additional control predictors to show that the correlations between my predictors and modification rates are significant. Tables 3.8 and 4.7 provide specific model results. In general, selectional strength had a simple effect with respect to both direct objects and instruments. Higher selectional strength predicted a lower chance of modification. That is, speakers were significantly more likely to modify a role filler (be it a direct object or instrument) in an event description when the role filler was unpredicted or unlikely.

For high diversity direct object verbs, this effect was strengthened as determined by a complex interaction between selectional strength and verb diversity. That is, event descriptions with high diversity verbs were more strongly biased in this direction than event descriptions with low diversity verbs. The high diversity of the verb can be thought of as giving the speaker more confidence in her assumptions of predictability.

The selectional constraint (η) of the event description's verb was a significant predictor only for direct object role filler modification. In these cases, I found a complex interaction between selectional constraint and selectional strength. Like with verb diversity, we can frame this effect in terms of boosting the simple selectional strength effect. Namely, event descriptions with strong selectional constraint verbs (e.g., *drink*) had more strongly biased modification rates for changes in selectional strength than their weak selectional constraint counterparts (e.g., *enter*). In other words, a predictable role filler has an increased likelihood of being left bare (unmodified) but a predictable role filler for a verb with generally predictable role fillers has an even *higher* increased likelihood of being left bare (unmodified). Like with the selectional strength and verb diversity interaction, this interaction of a verb's selectional constraint and selectional strength can be explained in terms of confidence. A strong selectional constraint and high selectional strength means that the verb/filler pair is not just highly predicted but highly predicted for a generally predictable verb.

Verb class (i.e., syntactic optionality for direct objects and semantic optionality for instruments) is a significant predictor of role filler modification in both chapters. For direct objects, I found a complex interaction between verb diversity and verb class. To understand this intuition, I need to frame the findings in terms of the simple, significant effect of verb diversity. Event descriptions with high diversity verbs have lower role filler modification rates. Verb class attenuates that effect. If the high diversity verb is also a syntactically obligatory direct object verb (e.g., *enter*), the simple effect goes away. That is, the effect of verb diversity on modification rates is only important when syntactic omission is a possible choice for the speaker.

For instruments, I found a simple effect of verb class. Event descriptions with semantically optional instrument verbs are more likely to contain a modified instrument role filler than their semantically obligatory counterparts.

My secondary findings concern the relationship between optionality and selectional constraint (η). In Chapter 3, I find that syntactically optional direct object verbs generally have stronger selectional constraints than their syntactically obligatory counterparts. Likewise, I find in Chapter 4 that semantically obligatory instrument verbs generally have stronger selectional constraints than their semantically optional counterparts. The parallel between the two findings is that the verb class theoretically predicted to have more information encoded about its role fillers has stronger selectional constraints (as measured by η) on average.

5.2 Implications for Language Production

In Chapter 1, I introduced two caricatures of speaker types: the solicitous speaker who maximizes context and the rational agent speaker who minimizes effort. The solicitous speaker provides additional information around those least predicted parts of his event descriptions. The less predicted a role filler is, the more likely it will be modified. This modification reflex can be explained by Firth's Distributional Hypothesis: you "know a word by the company it keeps" (Firth, 1968, pg. 179–180). Less predictable role fillers bring their own company along for event descriptions. That is, a speaker is more likely to provide additional, contextual details around a role filler when it otherwise would be unlikely or unpredicted in the event description. Hence, the Maximum Context Hypothesis predicts higher modification rates for those arguments if the solicitous speaker more accurately characterizes normal production.

In contrast, the rational agent speaker provides additional information for the most predictable parts of her event descriptions. The more predicted a role filler is, the more likely it will be modified. This modification reflex can be explained by Grice's Maxims: be brief when possible but always be informative. More predicted role fillers (on their own, unmodified) lower the average informativity of an event description. Modifying such an argument is one means for boosting that average informativity to be less in breach of Grice's Maxims. Hence, the Minimum Effort Hypothesis predicts higher modification rates for those arguments if the rational agent speaker is the dominant prototype.

I have used noun phrase realization as a means to investigate the relative dominance of the two speaker types. By looking at the context of event descriptions in which speakers could choose to modify an argument, we can better understand their choices and biases.

Understanding the production biases of these two speaker types is important for experimental work, model interpretation, and model generation. Experimental sentences (used indirectly for norming data or directly as items) should be as close to natural language production as possible. Whichever production style we assume to be the dominant style will determine the conditions under which participants will consider it normal to modify a role filler and when they will consider it marked or odd to modify a role filler. Experimental data yields different interpretations if the participant assumes the Maximum Context Hypothesis rather than the Minimum Effort Hypothesis. Computational and statistical models are likewise improved by explicit encoding of these natural biases.

The most significant of these biases is the inverse correlation between modification and predictability, as predicted by the Maximum Context Hypothesis. For instrument and direct object role fillers, less predictable fillers (given the verb) are more often modified. From this bias, we can conclude that the Maximum Context speaker style is more dominant than the Minimum Effort speaker style. Models that assume a Minimum Effort speaker will invert the true relationship between predictability and modification, yielding marked sentences when unmarked sentences are desired.

The second interesting bias is that syntactically optional arguments are generally more predictable than syntactically obligatory arguments. These results extend Resnik (1993)'s findings to a larger corpus and to a wider variety of verb types. This bias can be taken as evidence for the lexical encoding on the verb of the most likely role fillers (or their features) of its argument positions.[1] This bias would also imply that there is at least a

[1] An alternative analysis would argue that this bias is consistent with lexical encoding but not indicative of an underlying relationship.

difference in strength of connection between syntactically optional arguments of a verb and syntactically obligatory arguments of a verb. Are these connections actually categorically different between verb classes? How would the different types or strengths of connections impact production choices other than modification rate?

The third interesting bias is that semantically obligatory arguments are generally more predictable than semantically optional arguments. These results are a quantitative confirmation of Koenig et al. (2008)'s qualitative analysis of semantic obligatoriness and predictability of instruments. Like with the syntactic optionality bias, this bias is evidence for lexical connections between a verb and its semantically obligatory arguments, even when those arguments are syntactically optional.

5.3 Implications for Language Comprehension

My results from Chapters 3 and 4 have implications for both computational models of semantic relatedness and psycholinguistic models of processing. In Section 5.3.1, I provide a brief introduction to models of semantic relatedness. Then, I explain how my results inform model creation. I propose several extensions that could improve performance for smaller training sets and introduce new features to improve precision for larger training sets.

In Section 5.3.2, I give an overview of psycholinguistic models of processing. I cover three changes to how these models make predictions derived from my results. These models are often used to either predict or control for reading time, which in turn serves as a proxy for processing and integration time. In the second half of that section, I discuss how my results impact interpretation of reading time results and how my conclusions can be used to better control spurious effects in studies.

5.3.1 Computational Models of Semantic Relatedness

This section provides a brief overview of computational models of semantic relatedness and how they tie in with psychological reality. Two large classes of these computational models are context-based models (Deerwester et al., 1990; Lund et al., 1995; Kintsch, 2001; Jones & Mewhort, 2007) and syntax-augmented models (Lin, 1997, 1998a; Strzalkowski, 1999; Wiemer-Hastings, 2000; Padó & Lapata, 2003).

Context-based models are the simplest instantiation of Firth's Distributional Hypothesis: the more contexts two words share, the more related they are. Models sensitive to first-order co-occurrence track the syntagmatic relations of a word. They are focused on the question of what other words occur near the target word. Models sensitive to second-order co-occurrence track the paradigmatic relations of a word. They are focused on the question of what other words co-occur with the same words as the target word.

For both models, the primary concern is how to define what counts as a word's context. Here, again, models can be divided into two major groups: document-based models and window-based models. Take, as an example, the corpus in Figure 5.1 consisting of a common tongue twister. Document-based models would initially be concerned with

Figure 5.1: A Biscuit-Based Example Corpus

D_1: A Box of Biscuits
D_2: A Box of Mixed Biscuits
D_3: A Biscuit Mixer.

computing the co-occurrence of each lemma within a document, as shown in Table 5.1. The document boundary is an arbitrary, conceptual unit scope normally defined in terms of an article, a paragraph, or a sentence. In this corpus, each document (D_1–D_3) happens to be a sentence. These co-occurrence numbers could then be used to categorize lemmas that occurred in the same set of documents. The word *box* will have the strongest connec-

Table 5.1: A Document-based Co-Occurrence Matrix

	D_1	D_2	D_3
a	1	1	1
biscuit	1	1	1
box	1	1	0
of	1	1	0
mixed	0	1	0
mixer	0	0	1

tions to the word *of* because they occur in exactly the same documents. It will have the weakest connection to *mixer* because they occur in complementary sets of documents.

In contrast, window-based models would be concerned with finding all the lemmas that occurred within a fixed window-size on either side of our target word. In Table 5.2, I have assumed a window-size of one word in either direction. Here, *box* would have the strongest connections to *a* and *of*, because the latter two words fall within *box*'s window the most often.

Table 5.2: A First-Order Co-Occurrence Matrix with a Window-size of One

	a	biscuit	box	of	mixed	mixer
a	0	1	2	0	0	0
biscuit	1	0	0	1	1	1
box	2	0	0	2	0	0
of	0	1	2	0	1	0
mixed	0	1	0	1	0	0
mixer	0	1	0	0	0	0

Syntax-augmented models have generally been less successful than context-based models. The most successful of these models have used syntactic paths to determine a word's context (e.g., Padó & Lapata, 2003). Instead of using a linear window around a word to determine the context, syntactic relations determine the window. Distance becomes a measure of linguistic structural distance, rather than string distance.

More concretely, the word *of* is one word away from *biscuits* in D_1 but two words away in D_2 using a linear window. If we instead measure the distance in terms of dependency relations, *of* is one preposition/head-noun link away in both cases. The word *mixed*, in D_2, can be connected to *of* in two dependency links: from *mixed* to *biscuits* as a modifier and then from *biscuits* to *of* as the head of the prepositional phrase. Padó & Lapata (2007) tried to address the question of whether length or type of dependency links between "contextually close" words mattered. Namely, they created a syntax-augmented model that allowed for the weights on different participant roles to be counted differently and for the distance from end-to-end to be experimentally restricted. Their experiments found that only using dependency links of length one and treating all link types the same (i.e., where link type can be treated as similar to participant role type) performed much better than treating certain links preferentially according to their status in Keenan and Comrie's Obliqueness Hierarchy (1977).

I have shown in Chapter 3 that syntactic optionality is strongly correlated with argument predictability (see Resnik, 1993, for similar results). Thus, we should expect to see improved performance of a syntax-augmented model when the syntactic optionality of the verb/argument pair is taken into account. Basic syntax-augmented models (as described above) have not shown these expected performance improvements. Introducing syntactic optionality as a feature in syntax-augmented models is a novel variant that could prove more reliable. Functionally, we can introduce syntactic optionality in two ways during training. Below, I provide the basic framework for these two approaches to integrating syntactic optionality into syntax-augmented models.

First, we could more heavily weigh verb/argument links for syntactically optional arguments than for syntactically obligatory arguments. This differential weighting would effectively change the perceived (or learned) significance of *eat*-style verbs compared to *bring*-style verbs. That is, if the verb/direct object argument pairs <*eat*,*pizza*> and <*bring*,*pizza*> both occur equal times in the training corpus, <*eat*,*pizza*> would have stronger connections in the final model because connection strength would be increased by more for each repetition.

Second, we could augment the model in case of implicit occurrences of syntactically optional roles (i.e., a syntactically optional but semantically obligatory role is not overtly expressed in the event description). That is, first we need to learn the most likely role fillers for a particular verb/argument relation. Next, we re-analyze the training corpus, specifically looking for omitted instances of the syntactically optional roles of verbs. We then increment the co-occurrence frequencies for this verb to treat each implicit occurrence of the role as an additional occurrence of the most frequent role fillers for that role.

The intent of this additional implicit incrementation is to balance the relative strength of association for syntactically optional roles against those roles which are syntactically obligatory. We do not have to assume that an implicit occurrence is really psychologically equivalent to an explicit mention of the most frequent role(s). Instead, we are assuming that evoking the event will at least partially activate in the mind of the speaker and listener the role filler's concept or its general features if the concept is not actually known.

In this experiment, let us assume that *eat* and *bring* again occur equally frequently (n = 100) with the direct object role filler *pizza*. Further, *pizza* is the most frequent or predicted direct object role filler for both verbs. However, there are an additional 50 occurrences of *eat* with an implicit direct object. Because *bring* can never occur with an implicit direct object, there are zero instances for these types of verbs. In the strongest interpretation, the implicit occurrences would be treated as an additional 50 occurrences of <*eat*,*pizza*>. In a weaker interpretation, the implicit occurrences would be treated as additional occurrences of the various fillers according to their relative frequency.

Semantic optionality can likewise be integrated into model training. It can be used to weight arguments differently. That is, semantically obligatory role fillers (like the instruments in *stir* event descriptions) would be weighted more than semantically optional role fillers (like the instruments in *kill* event descriptions). Semantic optionality can also be used to boost verb co-occurrence frequencies for obligatory arguments with the most likely role fillers when no role filler is syntactically present. In this way, semantic optionality can be used identically to syntactic optionality for model training.

My next proposal for improving computational models leverages semantic relatedness. Roland et al. (2012) have shown that semantic neighborhood effects of possible role fillers are psychologically potent. According to their Semantic Similarity Hypothesis, processing is facilitated to the extent that the given role filler is semantically similar to other possible role fillers (pg. 268). Processing the instrument of verbs like *jab* (with many semantically similar role filers) is easier than processing the instrument of verbs like *attack* (with many dissimilar role fillers).

As a result of these findings alongside my own large corpus analysis of semantic optionality, we should be able to integrate semantic neighborhoods into my models of selectional strength and selectional constraint to further improve models. Like with the two implicit reference model boosts for syntactically optional and semantically obligatory role fillers, integrating semantic neighborhood effects require two passes through the training data.

In this two-step process, we first need to generate the semantic neighborhoods. The synset analysis that I used to calculate selectional strengths is one of many methods for generating semantic neighborhoods (for additional methods, see Deerwester et al., 1990; Lee, 1997; Sahlgren, 2006; Elman, 2009; Erk et al., 2010, among others). Then we can use the neighborhood network (per Roland et al.) to provide more context to the strength of connection between each verb/filler pair. The first pass generates semantic neighborhoods for each verb/role relation.[2] The second pass allows us to boost or diminish the strength of connection between a verb and role filler as a function of how strongly the role filler matches the dominant semantic neighborhood for that verb/role relation, thus matching the semantic neighborhood effect found by Roland et al.

The final improvement to training computational models of semantic relatedness is based on the correlation between modification and predictability. Namely, for both direct object and instrument arguments, less predicted role fillers are more likely to be modified. Computational models of semantic relatedness try to learn the first half of this correlation and have access to the second half. That is, in model training, we know whether the argument is modified or not. If we assume (as supported by my studies in Chapters 3 and 4) that modified role fillers are generally less expected, then we can use this correlation to affect weights learned during training. A simple training system would increment the strength of connection by one for each occurrence of a verb/role filler pair in the corpus. Taking the correlation into account, we would want to increment pairs with a modified role filler by 0.75 and pairs with a bare role filler by 1.25 (or some equivalent differential weighting).[3] This differential weighting should allow us to train a model on smaller corpora by bootstrapping predictability through speakers' modification choices. The capacity to use smaller corpora opens up this line of research to analysis in specialized sub-domains and analysis in languages with fewer resources.

[2]Using shared verb/role relations as the defining environment for neighborhood strength presupposes paradigmatic relationships are the most important for Roland et al.'s neighborhood effect. Syntagmatic relationships could just as easily be used to compute neighborhoods. Either relationship, or a third more complex definition of neighborhoods, can easily be used for the first pass.

[3]If the Minimum Effort Hypothesis had proven more resilient than the Maximum Context Hypothesis, these weights would be flipped.

5.3.2 Psycholinguistic Models of Processing

Like computational models of semantic relatedness, psycholinguistic models of processing can be improved by integrating the results of my studies into their calculations. First, I explain how my results relate to models of comprehension. Second, I explain how my results can impact the design and analysis of reading time and eye-tracking studies.

Several recent psycholinguistic models tie processing difficulty directly and indirectly to the conditional probability of all sentence continuations following the current word and how the current word matches or changes those probabilities (e.g., Jurafsky, 1996; Hale, 2001, 2006; Levy & Jaeger, 2007; Levy, 2008). I will specifically apply my result to Levy (2008)'s model of surprisal, but other models can similarly be augmented by my findings. Within Levy's model (and based on Hale, 2001), the difficulty of processing a word w_i is the negative log-probability of that word given its context, as shown in Equation 5.1.

$$difficulty \propto log\ P(\ w_i \mid w_{1...i-1}\ ,\ CONTEXT\ ) \tag{5.1}$$

The words $w_{1...i-1}$ denote the previous sentential context. The $CONTEXT$ denotes any extra-sentential context.

This formula, as a term for processing difficulty, is derived from the stipulation that the difficulty of processing a word is caused by updates to the listener's internal representation of the (probabilistically) preferred structural interpretation of the sentence. That is, fully comprehending a sentence requires internally representing its structure. Before a final structure is chosen, the listener creates a distribution of possible structures. For simplicity's sake, let us assume that the distribution of structures is updated after every new word until all the words in an utterance have been processed. Once all the words have been processed, the listener chooses the best structure given the context. Any word that creates a large change in the distribution of structures is said to be difficult. If the probability of a word w_i given the current context is low, then other, more likely words have been displaced in the listener's structural distribution.

My definition of a verb's η (or selectional constraint) aligns strongly with the same intuition of processing expectation in a surprisal-based model.[4] Low η values mean that the verb's role fillers are generally unpredictable. These types of verbs will do little to update the distribution of possible continuations for their role fillers. High η values mean that the verb's role fillers are generally predictable. These types of verbs can greatly update the distribution of possible continuations. As such, we should see lower surprisal values at the argument for verbs with low η values because those arguments will generally be sampled from a flatter distribution. No single argument or class of arguments dominate the listener's expectations. Without a strongly preferred continuation for the expected argument, there is not as large a change in expectation when a less preferred argument is chosen. This correlation could be useful as a descriptive factor in helping to understand how and why surprisal works. It could also be added as a component of the modeling itself, if η or an equivalent value is not already integrated into the calculation of a word's expectedness.

The second and third connections between my results and entropy-based models of expectedness would need to be integrated into the calculation of expectation and surprisal. I have shown that modification rates are higher for unexpected arguments. Thus, when a listener hears the beginning of an argument and that beginning is some type of modification, it may be reasonable for him to assume an otherwise less predicted argument head will follow. These expectations are similar to those explored by Jaeger (2005) in his study on *that* reduction. Jaeger found that speakers were more likely to include the

[4]Because surprisal is incrementally calculated, we need to consider what information is available when. Selectional strength, while a better parallel to processing expectation globally, is not available until after both the verb and filler have been produced. The verb's selectional constraints are available at the verb, and thus can impact surprisal sooner.

syntactically optional relativizer *that* when the following clause was less predicted. These types of cues could be used by listeners to ready themselves for unexpected continuations.

Likewise, a highly unexpected argument role filler that was not pre-modified could change the listener's expectations for post-modification. My analyses in Chapters 3 and 4 does not distinguish between pre- and post-modification rates. However, an early analysis that did distinguish between pre- and post-modification did not find any differences between the two places of modification with respect to expectedness. Thus, it stands to reason that an unexpected argument that had not yet been modified would still be a likely candidate for modification. This modification would have to be realized after the head noun.

Finally, my results are important to designing and interpreting on-line experiments. Hofmeister (2010) has shown that over-specification (that is, modifying a referent beyond the strict referential necessities) forces deeper semantic processing locally and can lead to lighter processing load later. That is, overspecification seems to convince readers to invest in deeper processing sooner. This deeper processing would have otherwise not been done or would have been done at a later stage in the sentence.

Since modification rates for unexpected arguments are significantly higher than for expected arguments, Hofmeister's results bring up the possibility of differential reading times between expected and unexpected arguments. The modified *and* expected arguments would be perceived as overspecified more often, which would result in deeper processing of expected arguments and lighter processing loads later in sentences containing expected arguments. One might argue that we should always expect to see different reading times for more predicted and less predicted argument role fillers, but these different reading times go against the natural bias. Namely, predicted arguments are expected to be easier to read and integrate than unpredicted arguments. If predicted arguments trigger deeper processing due to being overspecified, then reading times will slow down and a confound will be introduced in the design.

Engelhardt et al. (2006) have also shown negative effects of overspecification in reading studies. While listeners do not judge overspecified arguments as any worse than appropriately specified arguments, readers' eye movements indicate confusion. Again, if we assume artificially modified expected arguments are more likely to be seen as overspecified than their less expected counterparts, we will again see differential results for these two argument classes in contradiction to expected results.

As I have shown, my results are important both for computational models and psycholinguistic models. Training and development of these models can benefit from a better understanding of modification and how modification may be perceived as informative or as an indicator of overspecification. Overspecification has been shown by others to lead to differential processing and eye movements. If the effects of overspecification are ignored, researchers risk introducing confounds into experiments. If the effects of overspecification are taken into account, researchers can more carefully construct experimental items and norming data. Of more direct application, knowledge of the semantic optionality and syntactic optionality of a role filler given a verb can improve model training and provide tools for extrapolating from sparse datasets.

5.4 Future Work

Future work can be divided between correcting or refining previous work and extending work into new domains. Reducing errors in previous analyses will not change the overall tenor of my findings but will help to make any derived models more accurate and potentially viable with smaller training sets. After discussing several approaches to error reduction, I outline new directions for investigations.

5.4.1 Cleaning Up False Positives

In Section 2.4, I analyzed the types of errors introduced by my automated annotation system. There are experimental methods that I could introduce into the system that may reduce these errors. There are also alternate means for finding and annotating my input corpus that could result in lower error rates.

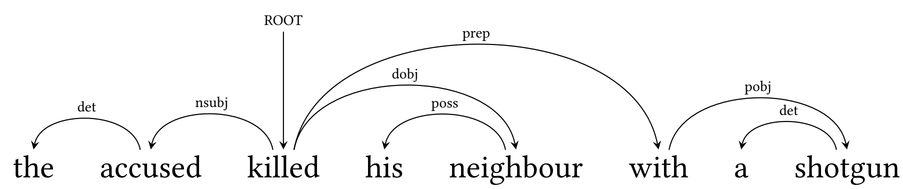
The first class of errors were from compound or collocated arguments that syntactically appear to be modified role fillers. I could use a list of common English compounds (or collocations) to filter out these entries. WordNet, for instance, lists frequent compounds. This filtering should hopefully reduce the number of falsely identified modified role fillers. However, this approach will potentially result in falsely identified bare role fillers. While the term *butcher's knife* seems more likely to be used as a compound noun, a normal knife owned by a butcher could be referred to using the same surface realization. A simple blacklist approach could not distinguish between these two referential intents. Likewise, this method would only have as good a coverage of possible phrases as determined by the input list.

Instead, I could use an automated approach to identifying compounds and collocations. Turney (2001)'s point-wise mutual information measures the bi-directional predictability between two units (e.g., words). This predictability targets the same underlying association I have tapped in using selectional strength. Syntactically modified role fillers with significantly high point-wise mutual information scores with their modifiers would be categorized as bare instead of modified (assuming that modifier was the only indicator of modification). One advantage of this approach over the previous approach is that it is more data-driven and requires less manual specification.

5.4.2 Alternate Methods for Matching Event Descriptions

The second class of errors arose from bad parsing. Some of these errors could be avoided through refinement of the `tgrep2` patterns in Appendix C. A likely more profitable change would be to use a more advanced parser (e.g., a retrained Charniak Parser) or a different style of parser. A dependency parser is a particularly good candidate as a replacement parser. A dependency parser (e.g., Lin, 1998b; de Marneffe et al., 2006) annotates the dependency relations explicitly instead of the tree-style annotations of a parser like the Charniak Parser. A sentence like Example 1a (repeated below) would be annotated as shown in Example 108. Every arc represents a dependency relation. Finding role fillers for a verb requires walking the arcs from the verb. The labels "bare" and "modified" would similarly require inspecting arcs out of the argument's head.

(1a′) the accused killed *his neighbour* **with a shotgun**

(108)


Some of the parsing errors could be mitigated by using an automated semantic role labeler (e.g., Simmons, 1973; Gildea & Jurafsky, 2000, 2002). These systems are trained to identify role relations based on a complex of syntactic, semantic, and lexical features. These labeled roles would largely replace the tree-matching described in Section 2.3.

5.4.3 More Complex Corpus Analysis

In addition to correcting old errors, there are many options for model refinement. Roland (2001) has shown that controlling for verb sense reduces cross-corpus subcategorization variation, improves norming data and experimental design, and could boost statistical parser performance. I have restricted myself above to analyzing verbs as if there is only a single possible optionality licensing status per role filler. That is, if a verb is categorized as syntactically optional with respect to its direct objects, all uses of the verb are treated the same. As a counter example, the verb *eat* is largely syntactically optional. The direct object is syntactically obligatory when used in the idiom 'to eat your own hat.' I could further refine my analysis by integrating verb sense into my measures. Thus, instead of calculating the η value (selectional constraint) for *cut*, I would calculate η for cut_1, cut_2, etc.

As I have treated each verb as a single conceptual reference, I treated modification as a homogeneous type. Sedivy and colleagues (Sedivy et al., 1999; Sedivy, 2003, 2006; Grodner & Sedivy, 2011) have experimentally found that readers are sensitive to physical property modifiers or percept modifiers like color and size (see also Altmann & Kamide, 1999; Arts et al., 2011; Horowitz & Frank, 2012, for studies from other groups). Readers actively use these semantic classes of modifiers to help with disambiguation before the head noun is known. Not all modifier classes were actively used by participants as cues. It may therefore be informative to analyze the semantic classes of modification in my corpus. In providing contextualizing modifiers for unlikely role fillers, speakers may prefer certain classes of modifiers to others.

A more general approach would be to extend the measure of selectional strength to also include role filler head/modifier pairs. That is, instead of treating all modifiers equally, highly predicted modifiers for a role filler may be included under different circumstances than less predicted modifiers. Highly predicted modifiers presumably provide better context for a role filler than less predicted modifiers. Instead, it could be that less predicted modifiers are used with less predicted role fillers.

Finally, I analyzed direct objects and instruments in separate models. In reality, all of these role filler expectations are computed in a single lexical system. A more realistic picture of an interlocutor's relative expectations could be derived from a single semantic network or neural network representation. The true predictability of a filler requires that we take into account all of its uses in all of its different semantic and syntactic roles. Including all the roles in a single model of predictability and relative entropy would also allow us to account for predictability contingencies between different role fillers (cf. Elman, 2009, for more relevant examples and a further discussion). For example, the most predictable instrument for *cut* depends a great deal on the direct object. Cutting something like a steak makes instruments like a knife more predicted whereas cutting something like a log makes instruments like a saw more likely (see, for instance, Matsuki et al., 2011)

5.4.4 On-Line Experiments

My results also lend themselves to several on-line experiments. Some of these, I covered in previous sections of this chapter. Another experiment is inspired by the continuation studies used by Rohde (2008) to investigate coherence effects between sentences. Rohde and colleagues (see Stevenson et al., 1994; Arnold, 2001; Rohde et al., 2006, 2007, for similar experiments) found that people's pronominal interpretations were biased by the event structure of the prior sentence. That is, passages like Example 109 biased interlocutors to continue the second sentence differently than passages like Example 110.

(109) John$_{SOURCE}$ handed a book to Bob$_{GOAL}$. He _____.

(110) John$_{SOURCE}$ was handing a book to Bob$_{GOAL}$. He _____.

The ambiguous pronoun for the first class of passages was more often resolved to be coreferential with the goal argument of the initial sentence whereas the second class of passages were biased towards a source argument solution.

Rohde's explanation for this pattern was tied to the different event structures. Example 109 is a completed event (as indicated by the perfective), which results in a focus on the end-state. As such, the goal argument will be more focused and, hence, more likely to be resolved as coreferential with the pronoun. In contrast, Example 110 is an incomplete event (as indicated by the imperfective), which cannot be focused on the end-state because it has not been achieved. As such, the source argument is salient to the ongoing event and, hence, more likely to be resolved as coreferential with the pronoun. Rohde et al. (2007) similarly looked at how discourse relations (specifically, coherence between sentences) could alter a reader's choice of ambiguous pronominal resolution.

With respect to semantically obligatory and semantically optional arguments, the semantically obligatory arguments have stronger connections to their verbs. Will this stronger underlying connection to the verb make them seem more salient[5] to interlocutors? Or will the surprise of an included semantically optional argument make those role fillers seem more salient?

To test these two different hypotheses, we can use a similar pronoun resolution continuation study to that described above. The two passages in Example 111 contrast only in their use of a semantically obligatory instrument verb (111a) or a semantically optional instrument verb (111b).

(111) a. Velma was <u>stirring</u> the soup with a ladle. It _____

b. Velma was <u>serving</u> the soup with a ladle. It _____

Comparing the percentage of continuations that resolve the ambiguous pronoun *it* to the *ladle* will tell us in which condition interlocutors feel that the instrument is more salient.

[5] I have chosen to focus on saliency in these studies because of its direct connection back to the Minimum Effort Hypothesis bias to elaborate on predicted role fillers and the Maximum Context Hypothesis bias to elaborate on the unpredicted role fillers. The relationship between predictability and discourse continuation is another avenue that could be pursued.

5.4.5 Predicting Syntactic Optionality

An argument's predictability cannot be a sufficient condition for licensing syntactic optionality. The verb's *devein* and *diagonalize* each have exactly one acceptable direct object role filler (as in Example 112) but neither license the syntactic omission of their direct object (as in Example 113).

(112) a. He deveined *the shrimp.*
b. She diagonalized *the matrix.*

(113) a. * He deveined.
b. * She diagonalized.

One possible argument for why the objects of these two verbs cannot be omitted is that they are both from a highly technical register. Because of their restricted use, these verbs have not undergone whatever normal process results in syntactic optionality. A restricted use argument is similar to the effects of verb diversity on a speaker's confidence in role filler predictability. High verb diversity means that the verb occurs in a lots of contexts and so any predictions should be fairly robust across contexts. Low verb diversity means that the verb occurs only in a small set of contexts, reducing a speaker's confidence in the global applicability of any learned patterns.

Another explanation for the obligatory status of these verb/argument pairs is that they are actually verb/direct object collocations. Thus, the direct object is not omissible because then that would break the collocational frame.

Neither of these arguments apply to a verb like *devour*, which also has a syntactically obligatory direct object (as in Example 114). It is not a highly technical term nor is its range of direct objects restricted to a reasonable collocational set. In fact, its range of direct object role fillers is very similar to that of the verb *eat*. Of course, the telicity may just be a consequence of how much we do or do not care about the specific details for cultural reasons.

(114) a. She devoured a pizza.

b. * She devoured.

Vendler (1967)'s *Aktionsart* provides one unifying characteristic for categorizing *devein*, *diagonalize*, and *devour* in contrast with strong selectional constraint, syntactically optional direct object verbs like *eat*. The verbs *devein*, *diagonalize*, and *devour* are all predominantly telic verbs.[6] A shrimp must be fully cleaned before it can be described as *devein*ed. The math must be completed before the *diagonalize*d matrix can be said to exist. The food must all be gone for it to have been *devour*ed. In contrast, verbs like *eat* and *steal* have strong telic and atelic senses.

One test for telicity is the appropriateness of the adverbial phrases "in an hour" (telic actions) or "for an hour" (atelic actions). Comparing the sentences in Example 115 with those in Example 116, we can see that *eat* clearly allows for both types of telicity.

(115) Telic

a. He ate dinner in an hour.

b. * He ate pizza in an hour.

c. He ate a pizza in an hour.

(116) Atelic

a. * He ate dinner for an hour[7].

b. He ate pizza for an hour.

c. * He ate a pizza for an hour.

The verb *devour*, however, only allows for telic interpretations, as in Example 117 versus Example 118.

(117) Telic

a. She devoured dinner in an hour.

[6] At the least, Mittwoch (1971, 1982), Browne (1971), Hopper & Thompson (1980), and Fillmore (1986) have also all put forth explanations tied to telicity or 'degree of transitivity' as explanations for the verbs that should license omission, due to argument predictability, but do not. In their present state, none of our combined explanations present a fully satisfactory solution.

[7] This is a delicate example. It does seem acceptable to say 'He spent an hour eating the dinnertime meal.' However, dinner **was** successfully eaten at the end of that hour. If the speaker consider dinner to be a specific amount of food (e.g., 'He only got through half his dinner', then this is a different sense of *dinner*.

b. * She devoured pizza in an hour.

c. She devoured a pizza in an hour.

(118) Atelic

a. * She devoured dinner for an hour.

b. * She devoured pizza for an hour.

c. * She devoured a pizza for an hour.

By default, the verbs *diagonalize* and *devein* do not allow for atelic interpretations. However, if the direct object is a bare plural (e.g., *matrices*) or a mass noun (e.g., *shrimp*), then an atelic interpretation can seemingly be achieved (see Examples 119–122).

(119) Telic

a. * He diagonalized matrices in an hour.

b. He diagonalized a matrix in an hour.

(120) Atelic

a. He diagonalized matrices for an hour.

b. * He diagonalized a matrix for an hour.

(121) Telic

a. * She deveined shrimp in an hour.

b. She deveined a shrimp in an hour.

(122) Atelic

a. She deveined shrimp for an hour.

b. * She deveined a shrimp for an hour.

This atelic event is actually an unknown quantity of telic events. The number of matrices successfully (and completely) diagonalized in Example 120a and the shrimp successfully (and completely) deveined in Example 122a is unknown. Nonetheless, the event described is an unbounded set of telic events. Rightly, these sentences should be read with the adverb "repeatedly/serially/sequentially(?)", as in Example 123. These events contrast with "eating for an hour" in that diagonalizing and deveining have binary results while eating can be done partially. In other words, "eating pizza for an hour (but not to completion)" still implies that some eating occurred. In contrast, "deveining shrimp for an hour (but not to completion)" implies failure.

(123) a. He repeatedly/serially/sequentially diagonalized matrices for an hour.

b. He repeatedly/serially/sequentially deveined shrimp for an hour.

Thus, *devein* and *diagonalize* appear to have atelic uses in these examples but are underlyingly always telic. Something about the semantics of a telic-only verb like *devour*, *diagonalize*, and *devein* may either block the process that allows a verb to become a syntactically optional verb or may force a verb to be syntactically obligatory (depending on which verb class you believe to be the default licensing). More verbs need to be analyzed according to this rubric in these examples to get a clearer understanding of the trends and to better understand what the exact relationship between telicity and syntactic obligatoriness might be.

A similar analysis to the telicity argument is Rappaport Hovav and Levin's result verb argument. In Rappaport Hovav & Levin (2001, pg. 779), they claim the Argument-per-Subevent Condition: an overt argument must exist for every subevent in the event structure. As a logical consequence, "verbs of change of state do not [allow unspecified objects]". The first crack in this analysis appears when we consider the distinction between manner verbs and result verbs, as per Rappaport Hovav & Levin (2010). Manner verbs (of which *devour* is a good example) do not force arguments to be overt like result verbs. Hence, we have lost our explanation for why the direct object of *devour* is syntactically obligatory.

Other problems with an easy analysis for *devour* arise when we consider the great polysemy of *eat*, our contrastive baseline for *devour*. Doug Roland (p.c.) makes the point that eating as a cultural event is very different from eating to achieve satiation. For the cultural event, we are more likely to care whether or not someone has eaten without regard to the specific food. For the satiation event, we are more likely to care about the specific food. The verb *devour* does not share the cultural event sense with *eat*. It is exclusively a manner of eating to achieve satiation. How much does having distinct senses (which focus on different frames, scripts, participants, etc.) influence the syntactic optionality of a verb? It may not be just the number of senses that distinguish *eat* and *devour*. Perhaps it is more important that *eat* has at least one sense that accommodates the event focus rather than the argument focus while *devour* has none. The underlying question here is not how *eat* and *devour* differ but how much of syntactic optionality is actually best explained by cultural biases or other extra-linguistic factors.

In conclusion, corpus patterns argue that speakers tend to augment those parts of an event description that are least broadly and/or locally predictable. These patterns should have implications for on-line processing by listeners (i.e., if listeners take modification as a cue of upcoming novelty) and production models of the mental lexicon (i.e., assuming a need to integrate both local and general predictability in the mental lexicon and planning stages). In either case, these results will help use create better language models through an improved understanding of why speakers may choose to modify arguments and how predictability can shape event descriptions.

Index of Word Lemmas

Appendix A

Verb Properties

Verb	Direct Object				Instrument			
	Syn. Opt.	η	BNC Count	Mod. %	Sem. Opt.	η	BNC Count	Mod. %
acquire					opt	-0.02	13	0.62
add					opt	0.16	16	0.50
align					opt	0.46	13	0.62
answer	opt	2.67	1155	0.33	opt	-0.36	11	0.64
approach	opt	-1.12	655	0.51	opt	-0.60	63	0.48
arrange					opt	0.09	11	0.64
attack					opt	-0.05	21	0.33
attend	opt	0.57	1095	0.67				
beat					obl	-0.30	29	0.28
become					opt	-0.64	77	0.51
begin					opt	-0.51	56	0.50
bombard					obl	0.27	13	0.23
break					opt	0.48	19	0.42
bring	obl	-0.61	2644	0.62	opt	-0.54	59	0.47
build	opt	-0.91	1030	0.64	obl	-0.32	27	0.63
buy					opt	-0.46	38	0.66
call	opt	0.29	2433	0.47	obl	0.88	6	
carry					opt	-0.34	17	0.35
catch	obl	0.61	1881	0.48	opt	-0.34	35	0.40
celebrate					opt	-0.65	36	0.42
change	opt	-0.90	3451	0.37	opt	-0.47	3	
charge					opt	-0.16	28	0.43
choose	opt	-1.28	1299	0.64	opt	-0.32	8	
close					opt	0.08	25	0.40
combine					opt	-0.64	77	0.57

Verb	Direct Object				Instrument			
	Syn. Opt.	η	BNC Count	Mod. %	Sem. Opt.	η	BNC Count	Mod. %
compare					opt	-0.90	109	0.54
complete					opt	0.02	21	0.57
conclude					opt	-0.50	25	0.40
confirm					opt	-0.21	18	0.50
connect					opt	0.03	33	0.45
console					opt	-0.30	16	0.38
contain					opt	-0.12	12	0.75
continue	opt	-0.07	969	0.57				
cover					obl	0.19	81	0.40
create					opt	0.04	25	0.44
cut	opt	-0.55	1308	0.50	obl	-0.77	10	
describe					opt	-0.63	12	0.58
do	obl	0.73	9047	0.51				
draw	opt	-0.90	1486	0.64	obl	-0.43	7	
drink	opt	2.14	645	0.43	opt	0.13	11	0.45
drive	opt	0.60	596	0.51	obl	-0.29	22	0.36
drop					opt	-0.60	12	0.50
eat	opt	1.34	1310	0.46	opt	0.14	7	
end					opt	-0.39	45	0.47
enter	opt	-1.14	2065	0.48	opt	-0.60	36	0.50
entertain					opt	-0.09	13	0.54
establish					opt	-0.41	14	0.50
examine					opt	-0.10	18	0.67
exchange					opt	0.16	14	0.21
explain	opt	-0.42	1126	0.68				
eye					opt	-1.09	22	0.32
fill					obl	-0.92	128	0.36
find	obl	-0.73	7476	0.65	opt	-0.97	132	0.51
finish					opt	0.17	18	0.56
fix					opt	-0.43	22	0.55
fly	opt	0.48	269	0.45	opt	0.50	3	
follow	opt	-1.34	3299	0.70	opt	-0.60	75	0.57
form					opt	0.10	15	0.47
gain	opt	-0.42	1208	0.74				
get	obl	-0.48	20106	0.51	opt	-0.62	94	0.46
give	obl	-0.25	9069	0.69				
grab	opt	0.16	494	0.36	opt	0.78	5	
handle					opt	-0.28	11	0.55

Verb	Direct Object				Instrument			
	Syn. Opt.	η	BNC Count	Mod. %	Sem. Opt.	η	BNC Count	Mod. %
hang	obl	-0.14	154	0.44	opt	-5.57	1	
have	obl	-0.06	46966	0.72				
hear	opt	0.36	3386	0.57	opt	0.83	3	
help					opt	-0.65	51	0.24
hit	obl	-1.10	1200	0.39	obl	-0.51	97	0.33
hold					opt	-0.54	64	0.53
identify					opt	-0.50	32	0.38
illustrate					obl	0.07	20	0.50
include					opt	0.01	20	0.40
increase					opt	-0.46	20	0.80
indicate					opt	0.61	15	0.80
join					opt	-0.55	39	0.54
judge	opt	-0.17	200	0.60				
keep					opt	-0.52	27	0.41
kick	opt	0.14	256	0.41	opt	-0.37	17	0.59
kill					opt	-0.28	24	0.50
know	opt	-1.40	4302	0.46				
launch					opt	-0.08	30	0.53
lead	opt	-0.42	831	0.64	opt	0.49	21	0.76
leave	opt	-0.36	5390	0.39	opt	-1.39	303	0.53
lift					opt	0.12	18	0.61
like	obl	0.14	3552	0.44				
link					opt	-0.68	94	0.47
look					opt	-0.04	24	0.46
lose	opt	-1.20	3963	0.42	opt	-0.43	27	0.52
make	obl	-0.30	21626	0.58	opt	-1.33	277	0.41
mark					obl	-0.10	18	0.56
marry	opt	2.62	521	0.68				
meet					opt	-0.64	52	0.54
mix					opt	-0.06	17	0.65
mop					obl	-0.99	15	0.33
obey	opt	0.49	147	0.46				
open	obl	1.28	3387	0.34	opt	-0.70	60	0.58
order	opt	-0.36	531	0.68	opt	0.13	4	
pack	opt	1.02	215	0.28	opt	-0.12	6	
paint	opt	-0.47	287	0.54	obl	0.28	10	
pass	opt	-0.81	1251	0.57	opt	-0.51	46	0.46
pay	opt	1.50	2731	0.59				

Verb	Direct Object				Instrument			
	Syn. Opt.	η	BNC Count	Mod. %	Sem. Opt.	η	BNC Count	Mod. %
pick					obl	-0.48	11	0.45
play	opt	0.58	2911	0.51	opt	-0.34	69	0.39
pour	opt	2.03	314	0.44				
present					opt	-1.20	190	0.53
press					opt	-0.78	11	0.45
print	opt	0.15	193	0.63	obl	0.74	2	
prod					obl	0.38	11	0.27
produce					opt	-0.25	42	0.67
prove					opt	-0.23	15	0.73
pull	opt	-0.58	949	0.41	obl	0.22	4	
push	opt	-1.08	731	0.41	obl	0.37	12	0.33
put	obl	-1.58	3459	0.58	opt	-0.88	107	0.51
raise					opt	-0.47	12	0.17
reach					opt	-0.57	32	0.53
read	opt	-0.21	2361	0.42	opt	-0.64	26	0.50
recall	opt	-0.90	441	0.73				
receive					opt	-0.48	36	0.53
reflect					opt	-0.51	14	0.21
refuse	opt	0.65	443	0.57				
remember	opt	-0.60	1717	0.62				
replace					opt	-0.57	57	0.49
report					opt	-0.22	22	0.59
represent					opt	0.05	13	0.46
return					opt	0	13	0.54
reward					opt	-0.90	11	0.45
rub					opt	-0.11	36	0.33
saw	obl	-0.92	2587	0.67				
say	obl	1.21	9002	0.31				
score					opt	-0.32	23	0.74
see	obl	-0.45	8658	0.60	opt	-0.74	81	0.44
send					opt	-0.37	51	0.41
set					opt	-0.84	68	0.57
shake					opt	-0.63	26	0.27
show	obl	-0.46	4405	0.79	opt	-0.89	116	0.48
sing	opt	1.34	351	0.48				
speak					opt	-0.23	16	0.56
spray					opt	-0.25	11	0.27
stand					opt	-0.36	33	0.30

Verb	Direct Object				Instrument			
	Syn. Opt.	η	BNC Count	Mod. %	Sem. Opt.	η	BNC Count	Mod. %
start					opt	-0.57	61	0.54
steal	opt	0.20	400	0.43	opt	0.29	8	
strike					obl	-0.05	33	0.33
stroke					obl	-0.16	18	0.50
study					opt	-0.47	16	0.44
supply					opt	-0.56	47	0.53
support					opt	-0.49	25	0.72
swallow	opt	0.75	193	0.42				
sweep					obl	0.30	11	0.82
take	obl	-1.10	19394	0.54	opt	-1.49	297	0.47
tell					opt	-1.06	68	0.57
think	opt	-1.67	1492	0.55				
threaten					opt	-0.67	30	0.23
touch					opt	-0.43	22	0.32
treat					opt	-0.84	75	0.52
turn					opt	-0.38	24	0.54
view					opt	-1.14	34	0.29
want	obl	0.03	3913	0.47				
waste	opt	1.83	521	0.38				
watch	opt	-0.44	1862	0.46	opt	-0.77	61	0.46
wave	opt	0.96	265	0.42				
wear	obl	1.18	2063	0.62				
wipe					obl	-0.29	12	0.17
write	opt	-0.14	2086	0.58	obl	-0.15	6	

The author intended for this page to be blank.

Appendix B

Working Through Relative Entropy Examples

Let us assume a rather minimal and repetitive corpus of English, as shown in Table B.1. There are only nine unique sentences, three verb lemmas (*drink*, *grow*, and *ferment*), and six attested direct object role filler lemmas. Each sentence occurs between 2 and 70 times.

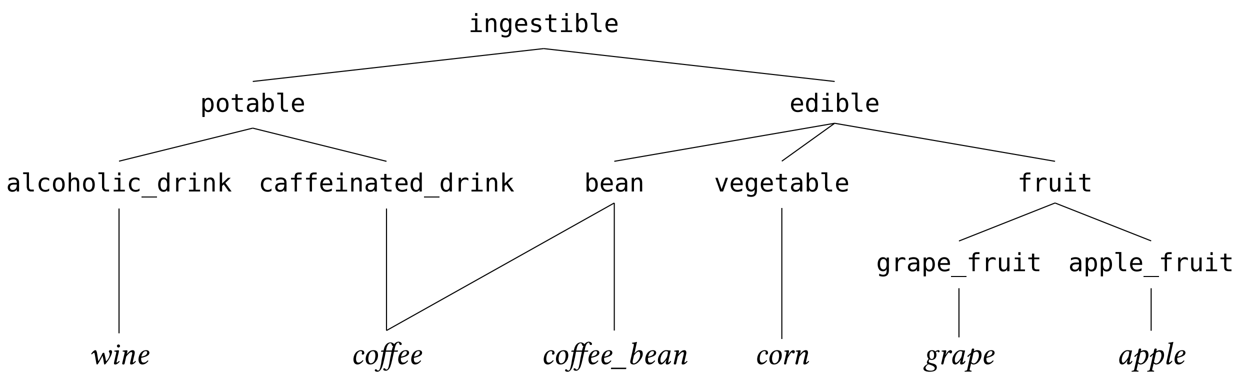
	Sentence	Count
1	She drank wine.	50
2	He drank coffee.	50
3	She fermented wine.	30
4	He fermented grapes.	70
5	She grew coffee.	2
6	He grew coffee beans.	13
7	She grew corn.	15
8	He grew grapes.	35
9	She grew apples.	35

Table B.1: Full Corpus Frequency Counts

Note that the verb *drink* has a balance between its two direct object fillers. In contrast, the verb *ferment* has a strong bias towards one of its fillers. The verb *grow* has a wider range of filler frequencies but there are still two fillers (*apple* and *grape*) that tend to dominate the counts.

I have also created a simplified semantic hierarchy that encodes the relations between the role filler concepts. I am treating these artificial concepts as correlates to WordNet synsets used in my dissertation (Fellbaum, 1998). In Figure B.1, you can see the six lemmas and their network of synset hypernyms. Use of the term *wine* triggers or activates three synsets: `alcoholic_drink`, `potable`, and `ingestible`. The term *coffee* is ambiguous between the drink and the bean. As such, its use triggers both conceptual chains:

Figure B.1: A Simple Hierarchy of Lemmas and Synsets



`caffeinated_drink`, `potable`, `ingestible`, `bean`, and `edible`. Even though `ingestible` is accessible through both conceptual chains, the concept is treated as only having been triggered once per use of *coffee* in an event description.

Combining the event descriptions in Table B.1 and the semantic hierarchy in Figure B.1, we can create a table of synset activation counts across the corpus and per verb. For instance, the term *wine* occurs 50 times as the direct object role filler in a *drink* event description and 30 times following *ferment*. Each use of the term *wine* counts as a mention of the synsets `alcoholic_drink`, `potable`, and `ingestible`. These numbers are reflected in the first column of Tables B.2–B.5. Table B.2 shows synset activation counts for the full corpus. As such, it is the sum of the three individual verb tables (i.e., Table B.3 for *drink* event descriptions, Table B.4 for *ferment* event descriptions, and Table B.5 for *grow* event descriptions).

	wine	coffee	coffee_bean	corn	grape	apple
ingestible	80	52	13	15	105	35
potable	80	52	0	0	0	0
edible	0	52	13	15	105	35
alcoholic_drink	80	0	0	0	0	0
caffeinated_drink	0	52	0	0	0	0
bean	0	52	13	0	0	0
vegetable	0	0	0	15	0	0
fruit	0	0	0	0	105	35
grape_fruit	0	0	0	0	105	0
apple_fruit	0	0	0	0	0	35

Table B.2: Full Corpus Frequency Counts

	wine	coffee	coffee_bean	corn	grape	apple
ingestible	50	50	0	0	0	0
potable	50	50	0	0	0	0
edible	0	50	0	0	0	0
alcoholic_drink	50	0	0	0	0	0
caffeinated_drink	0	50	0	0	0	0
bean	0	50	0	0	0	0
vegetable	0	0	0	0	0	0
fruit	0	0	0	0	0	0
grape_fruit	0	0	0	0	0	0
apple_fruit	0	0	0	0	0	0

Table B.3: 'drink' Event Frequency Counts

	wine	coffee	coffee_bean	corn	grape	apple
ingestible	30	0	0	0	70	0
potable	30	0	0	0	0	0
edible	0	0	0	0	70	0
alcoholic_drink	30	0	0	0	0	0
caffeinated_drink	0	0	0	0	0	0
bean	0	0	0	0	0	0
vegetable	0	0	0	0	0	0
fruit	0	0	0	0	70	0
grape_fruit	0	0	0	0	70	0
apple_fruit	0	0	0	0	0	0

Table B.4: 'ferment' Event Frequency Counts

	wine	coffee	coffee_bean	corn	grape	apple
ingestible	0	2	13	15	35	35
potable	0	2	0	0	0	0
edible	0	2	13	15	35	35
alcoholic_drink	0	0	0	0	0	0
caffeinated_drink	0	2	0	0	0	0
bean	0	2	13	0	0	0
vegetable	0	0	0	15	0	0
fruit	0	0	0	0	35	35
grape_fruit	0	0	0	0	35	0
apple_fruit	0	0	0	0	0	35

Table B.5: 'grow' Event Frequency Counts

The relative probability of activation of each synset has been calculated in Tables B.6–B.9. To make reading these and further tables easier, I have left empty those cells which should contain a zero. The denominator for these probabilities could possibly be determined by several external factors. I have chosen (following Resnik, 1993, and based on analyses discussed in his Appendix A) to define the probability such that the denominator is a unit credit assignment to the synset and the numerator is the sum of all synset counts. By unit credit assignment, I mean that an occurrence of a term like *wine* equally boosts each related synset by 1. The unit measure breaks the strict assumption of probability because the probability mass across the corpus no longer sums to 1. As a result of the change in aggregate probability mass, it also makes sense to change the numerator to the total mass of synsets rather than the total mass of event descriptions. That is, the 80 occurrences of *wine*/`ingestible` is evaluated out of a total of 1144 synset occurrences rather than out of the 300 event descriptions. The probability of *wine*/`ingestible` given the verb *drink* is 80/1144. See Equation B.1, below, for a precise equation followed by real values.

	wine	coffee	coffee_bean	corn	grape	apple	Aggregate
ingestible	7.0	4.5	1.1	1.3	9.2	3.1	26.2
potable	7.0	4.5					11.5
edible		4.5	1.1	1.3	9.2	3.1	19.2
alcoholic_drink	7.0						7.0
caffeinated_drink		4.5					4.5
bean		4.5	1.1				5.7
vegetable				1.3			1.3
fruit					9.2	3.1	12.2
grape_fruit					9.2		9.2
apple_fruit						3.1	3.1

Table B.6: Full Corpus Probabilities

	wine	coffee	coffee_bean	corn	grape	apple	Aggregate
ingestible	12.5	12.5					25.0
potable	12.5	12.5					25.0
edible		12.5					12.5
alcoholic_drink	12.5						12.5
caffeinated_drink		12.5					12.5
bean		12.5					12.5
vegetable							
fruit							
grape_fruit							
apple_fruit							

Table B.7: 'drink' Event Probabilities

	wine	coffee	coffee_bean	corn	grape	apple	Aggregate
ingestible	8.1				18.9		27.0
potable	8.1						8.1
edible					18.9		18.9
alcoholic_drink	8.1						8.1
caffeinated_drink							
bean							
vegetable							
fruit					18.9		18.9
grape_fruit					18.9		18.9
apple_fruit							

Table B.8: 'ferment' Event Probabilities

	wine	coffee	coffee_bean	corn	grape	apple	Aggregate
ingestible		0.5	3.5	4.0	9.4	9.4	26.7
potable		0.5					0.5
edible		0.5	3.5	4.0	9.4	9.4	26.7
alcoholic_drink							
caffeinated_drink		0.5					0.5
bean		0.5	3.5				4.0
vegetable				4.0			4.0
fruit					9.4	9.4	18.7
grape_fruit					9.4		9.4
apple_fruit						9.4	9.4

Table B.9: 'grow' Event Probabilities

Our primary concern for calculating selectional strength and selectional constraint is the final column of Tables B.6–B.9. This column contains the aggregate probability of each synset across the entire corpus (as in Table B.6) and by verb (as in Tables B.7–B.9).

As you may recall from Section 1.2.1, the relative entropy between a synset and a verb is determined by Equation 1.7, repeated below. We have two events ($p(x)$ and $q(x)$) with their own respective probabilities. We want to measure how knowledge of the second event ($q(x)$) changes our expectations about the first event ($p(x)$).

$$D(p||q) = \sum_{x} p(x) log \frac{p(x)}{q(x)} \tag{1.7'}$$

For the purposes of this dissertation, the equation can be simplified and some of the variables renamed for clarity, as in Equation B.1. Namely, we are interested in how our expectation about the likelihood of a concept occurring as a role filler ($p(c)$) changes if we know the verb used ($p(c|v)$).

$$p(c) log \frac{p(c)}{p(c|v)} \tag{B.1}$$

Looking at more concrete examples, the $p(c)$ is the probability of a role filler across the entire corpus (e.g., $p(potable) = 11.5\%$ and $p(alcoholic\_drink) = 7\%$). The $p(c|v)$ is the probability of a role filler given a particular verb (e.g., $p(potable|drink) = 25\%$ and $p(potable|grow) = 0.5\%$). The relative entropy for every role filler lemma/synset pair is provided in Tables B.10–B.12.

	wine	coffee	coffee_bean	corn	grape	apple
ingestible	-0.013	-0.013				
potable	0.089	0.089				
edible		-0.083				
alcoholic_drink	0.041					
caffeinated_drink		0.046				
bean		0.045				
vegetable						
fruit						
grape_fruit						
apple_fruit						

Table B.10: 'drink' Event Selectional Strengths

These calculations should make clear that selectional strength is not strictly relative entropy. Each role filler lemma can be associated with multiple relative entropy values (i.e., one for all of its related synsets). The selectional strength is the **strongest** of all of these relative entropy measures. For the verb/role filler pair of *drink* and *wine*, the selectional strength is `0.0892` and derives from the relative entropy for `potable` given

	wine	coffee	coffee_bean	corn	grape	apple
ingestible	0.008				0.008	
potable	-0.041					
edible					-0.003	
alcoholic_drink	0.010					
caffeinated_drink						
bean						
vegetable						
fruit					0.053	
grape_fruit					0.066	
apple_fruit						

Table B.11: 'ferment' Event Selectional Strengths

	wine	coffee	coffee_bean	corn	grape	apple
ingestible		0.005	0.005	0.005	0.005	0.005
potable		-0.354				
edible		0.063	0.063	0.063	0.063	0.063
alcoholic_drink						
caffeinated_drink		-0.097				
bean		-0.020	-0.020			
vegetable				0.015		
fruit					0.052	0.052
grape_fruit					0.002	
apple_fruit						0.034

Table B.12: 'grow' Event Selectional Strengths

drink.

Analyzing the relative entropy values for the verb *give*, you can see that the ambiguity between $coffee_{drink}$ and $coffee_{bean}$ has been automatically resolved by the pull of the other edible concepts that co-occur with the verb *grow*. Namely, the relative entropy for drink related synsets is low while the relative entropy for bean related synsets is high. The synset `ingestible`, which is a hypernym of both senses, has a relatively low score because it occurs across the board.

Finally, the η value (selectional constraint) for a verb is defined as the sum of all selectional strengths for that verb. To tabulate η, we sum the highest relative entropy values in each column for Tables B.10–B.12. The verb *drink* has a cumulative η value of 0.178. The verb *ferment* has a cumulative η value of 0.077. The verb *grow* has a cumulative η value of 0.317.

These numbers tell us that the verb *ferment*, with the lowest selectional constraint, has the least predictable set of direct object role fillers of the three verbs. It may appear from the event descriptions counts in Table B.1 that the role filler *grape* should be an obvious predictable choice for the direct object of *ferment*, but that intuition ignores both the use of *grape* with other verbs and the other role fillers for the verb *ferment*. First, the word *grape* is a relatively frequent role filler for *grow*, which means that the use of *grape* with *ferment* may seem high, but that is partially because *grape* is a highly frequent role filler in general. Second, *wine* (the other direct object role filler for *ferment*) and *grape* only share a synset hypernym at the root of the semantic hierarchy. There is no more precise generalization that can be made to bridge the lexically different choices at a semantic level.

In contrast, the verb *drink*, which appears to have a strict balance between role fillers, actually has a higher selectional constraint than *ferment*. This higher selectional constraint can be explained by the fact that *wine* and *coffee* share a synset (`potable`) at a lower level than the root of the semantic hierarchy. Thus, the system can effectively make a generalization about the types of role fillers that follow *drink* that is more precise than that the role filler must be a hyponym of `ingestible`.

The verb *grow* has the highest η value of all three verbs for two reasons. First, there is a non-trivial synset that encompasses all of the direct object role fillers (i.e., `edible` rather than the trivial `ingestible`). This generalized concept best typifies the role fillers of only this verb. Second, the synset does not reliably occur with other verbs. That is, it does occur with both *ferment* and *drink*, but it is not the convergence point for any direct object role fillers for either of these verbs.

Appendix C

tgrep2 Patterns

C.1 Finding Event Descriptions

I used the tgrep2 utility[1] and patterns adapted from a corpus-general structural frequency study described by Roland et al. (2007) to extract event descriptions with appropriate syntactic structures. The variable $VERB (in the patterns below) refers to an appropriate verb from the list in Appendix A. The other two variables ($NOT_PASSIVE and $BE_GET) are defined in Example 124.

(124) tgrep2 Variables

a.
```
BE_GET=is|are|was|were|be|am|been|get|gets|got|gotten|getting|being
```

b.
```
NOT_PASSIVE=!>(VBN>(VP!<PRT!>>NP%(AUX<<$BE_GET)))!>(VBN>(VP!<PRT!>>NP%(/VB/<$BE_GET)))!>(VBN>(VP>(VP!<PRT!>>NP%(/VB/<$BE_GET))))!>(VBN>(VP!<PRT!>>NP%(VP<(/VB/<$BE_GET))))
```

C.1.1 Direct Object Events

Five patterns for simple transitive sentences were also adapted from Roland et al. (2007). These patterns are listed in Example 125.

(125) tgrep2 Direct Object Verb Patterns

a.
```
TRANS1 (/^VP/%/^NP/=subject>>(`TOP|S1)!<PRT!>>/^NP/!<(/^NP/</NONE/)<(`/VB/|/AUX/<(($VERB=verb)$NOT_PASSIVE))<(/^NP/=object%..(/^S/!<TO!<(/^AUX/|/^VP/<TO)!<(/^VP/<VBG)))!<(/^NP/%../^NP/|/SBAR/|/^VP/|X))
```

b.
```
TRANS2 (/^VP/%/^NP/=subject>>(`TOP|S1)!<PRT!>>/^NP/!<(/^NP/</NONE/)<(`/VB/|/AUX/<(($VERB=verb)$NOT_PASSIVE))</^NP/=object!<(/^NP/%../^NP/|/PP/|S|/S-/|/SBAR/|/^VP/|X)!>(/^VP/>(S|SINV%S)))
```

c.
```
TRANS3 (/^VP/%/^NP/=subject>>(`TOP|S1)!<PRT!>>/^NP/!<(/^NP/</NONE/)<(`/VB/|/AUX/<(($VERB=verb)$NOT_PASSIVE))<(/^NP/=object%..(/^S/!<TO!<
```

[1] http://tedlab.mit.edu/~dr/TGrep2/

```
(/^AUX/|/^VP/<TO)<(/^VP/<VBG)))!<(/^NP/%../^NP/|/SBAR/|/^VP/|X))
```

d.
```
TRANS4 (/^VP/%/^NP/=subject>>(`TOP|S1)!<PRT!>>/^NP/!<(/^NP/</NONE/)<
(`/VB/|/AUX/<(($VERB=verb)$NOT_PASSIVE))<(/^NP/=object%..(/SBAR/!<
(/SBAR/<(IN<<(that)))!<(/WH/)!<(-NONE-)!<(IN)!<(/^S/!%__)))!<(/^NP/%..
/^NP/|S|/S-/|/^VP/|X))
```

e.
```
TRANS5 (/^VP/%/^NP/=subject>>(`TOP|S1)!<PRT!>>/^NP/!<(/^NP/</NONE/)<
(`/VB/|/AUX/<(($VERB=verb)$NOT_PASSIVE))</^NP/=object!<(/^NP/%../^NP/|
/PP/|S|/S-/|/SBAR/|/^VP/|X)>(/^VP/>(S|SINV%S)))
```

C.1.2 Instrument Events

Example 126 contains two basic patterns pertinent to instrument verbs. The first pattern matches those sentences with a prepositional phrase headed by *with*. The second pattern matches those sentences without a prepositional phrase. Because my primary concern is with instrument mentions rather than null instruments, I have only analyzed data which match the former pattern.

(126) tgrep2 Instrument Verb Patterns

a.
```
INSTRUMENT1 /^VP/%/^NP/=subject!>/^VP/<(`/VB/|/AUX/<(($VERB=verb)
$NOT_PASSIVE))<(/^NP/=object%..(/PP/<<(IN<<with)<</^NP/=instrument))
```

b.
```
INSTRUMENT2 /^VP/%/^NP/=subject!>/^VP/<(`/VB/|/AUX/<(($VERB=verb)
$NOT_PASSIVE))<(/^NP/=object!%..(/PP/<<(IN<<with)<</^NP/))
```

C.2 Labeling Bare and Modified Role Fillers

In order to verify that my patterns fully spanned the possible patterns, I started by counting the role fillers in bins based on the number of daughters present. Example 127 shows these patterns. No matched role fillers had more than eighteen daughters.

(127) Counting Daughter Nodes

```
c.1-daughter     NP!>>NP<(__!$.__!$,__)
c.2-daughters    NP!>>NP<(__$.(__!$.__)!$,__)
c.3-daughters    NP!>>NP<(__$.(__$.(__!$.__))!$,__)
c.4-daughters    NP!>>NP<(__$.(__$.(__$.(__!$.__)))!$,__)
c.5-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__!$.__))))!$,__)
c.6-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__!$.__)))))
                 !$,__)
c.7-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__$.(__!$.__
                 ))))))!$,__)
c.8-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                 !$.__)))))))!$,__)
```

```
c.9-daughters     NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                  $.(__!$.__))))))))!$,__)
c.10-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                  $.(__$.(__!$.__)))))))))!$,__)
c.11-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                  $.(__$.(__$.(__!$.__))))))))))!$,__)
c.12-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                  $.(__$.(__$.(__$.(__!$.__)))))))))))!$,__)
c.13-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                  $.(__$.(__$.(__$.(__$.(__!$.__))))))))))))!$,__)
c.14-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                  $.(__$.(__$.(__$.(__$.(__$.(__!$.__)))))))))))))
                  !$,__)
c.15-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                  $.(__$.(__$.(__$.(__$.(__$.(__$.(__!$.__))))))))
                  ))))))!$,__)
c.16-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                  $.(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__!$.__)))
                  ))))))))))))!$,__)
c.17-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                  $.(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                  !$.__))))))))))))))))!$,__)
c.18-daughters    NP!>>NP<(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                  $.(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__$.(__
                  $.(__!$.__)))))))))))))))))!$,__)
```

Most labeling could be achieved by looking only one level deep in the tree. As such, the nodes immediately underneath the head NP were usually sufficient. Patterns in Examples 128, 129, 131, and 132 matched most of the available phrases at this first depth of processing. However, some patterns included a noun phrase opaquely nested within a noun phrase at the first level and no other evidence clearly indicated that this role filler was modified. For these structures, I employed a two-stage processing such that the first pass pulled out the nested NPs, according to the patterns in Examples 130 and 133. These noun phrases were then matched again against the top level patterns in Examples 128, 129, 131, and 132. I have separated the direct object and instrument role filler patterns because of a few idiosyncrasies in the necessary patterns for complete coverage. A union of the two sets would provide the same total coverage but would have made analysis and bug testing more difficult.

C.2.1 Labeling Direct Object Role Fillers

(128) Bare Role Fillers

```
b.bare              NP!>>NP<(NN|NNS|CD<(__=dep)!$.__!$,__)
b.DT-only           NP!>>NP<(DT|PDT|/^PRP/|/^RB/|INTJ$.(NN|NNS|CD
                    <(__=dep)!$.__)!$,__)
b.IN-that           NP!>>NP<(IN<that$.(NN|NNS|CD<(__=dep)!$.__)!$,__)
b.wh-words          NP!>>NP<(WDT|WP$.(NN|NNS|CD<(__=dep)!$.__)!$,__)
b.other-one-mod     NP!>>NP<(!DT|CD|/^PRP/|PDT|QP|/^RB/|INTJ|IN|"$"|
                    ADVP|PRN|CC|","|WDT|WP|/^JJ/|/^NN/|NP|/^VB/|ADJP|
                    UCP$.(NN|NNS|CD<(__=dep)!$.__)!$,__)
b.PDT-DT            NP!>>NP<(PDT$.(DT<(__=dep)!$.__)!$,__)
b.NN-RB             NP!>>NP<(NN|NNS<(__=dep)$.(RB!$.__)!$,__)
b.DT-VBN            NP!>>NP<(DT|/^PRP/$.(VBN<(__=dep)!$.__)!$,__)
b.DT-pre-mods       NP!>>NP<(PDT|DT|INTJ$.(DT$.(NN|NNS|CD<(__=dep)
                    !$.__))!$,__)
b.RB-pre-mods       NP!>>NP<(RB$.(DT$.(NN|NNS|CD<(__=dep)!$.__))!$,__)
```

(129) Modified Role Fillers

```
m.Noun-PRN          NP!>>NP<(NNP<(__=dep)$.(PRN!$.__)!$,__)
m.CD-only           NP!>>NP<(CD|QP$.(NN|NNS|CD<(__=dep)!$.__)!$,__)
m.JJ-only           NP!>>NP<(/^JJ/|/^NN/|NP|/^VB/|ADJP$.(NN|NNS|CD
                    <(__=dep)!$.__)!$,__)
m.UCP-only          NP!>>NP<(UCP$.(NN|NNS|CD<(__=dep)!$.__)!$,__)
m.NN-PP             NP!>>NP<(NN|NNS<(__=dep)$.(PP!$.__)!$,__)
m.NN-QP             NP!>>NP<(NN<(__=dep)$.(QP!$.__)!$,__)
m.mod-VBN           NP!>>NP<(CD|NN$.(VBN<(__=dep)!$.__)!$,__)
m.leaf-postmod-1    NP!>>NP<(NN|NNS|VB|EX|JJ<(__=dep)$.(SBAR!$.__)
                    !$,__)
m.leaf-postmod-3    NP!>>NP<(NN|NNS|VBZ<(__=dep)$.(S!$.__)!$,__)
m.compound-NP       NP!>>NP<(NP<(/^NN/|JJ|JJS|PP|CD|QP!$.__)$.(NP|
                    NX=np!$.__)!$,__)
m.NP-PP             NP!>>NP<(NP=np$.(PP!$.__)!$,__)
m.NP-WHPP           NP!>>NP<(NP=np$.(WHPP!$.__)!$,__)
m.NP-post-mod-1     NP!>>NP<(NP=np$.(SBAR!$.__)!$,__)
m.NP-post-mod-2     NP!>>NP<(NP=np$.(VP!$.__)!$,__)
m.NP-post-mod-3     NP!>>NP<(NP=np$.(S!$.__)!$,__)
m.JJ-pre-mods       NP!>>NP<(__$.(ADJP|JJ|JJR|JJS|NN|NNP|VBN$.(NN|NNS|CD
                    <(__=dep)!$.__))!$,__)
m.JJ-pre-mods       NP!>>NP<(ADJP$.(DT$.(NN|NNS|CD<(__=dep)!$.__))!$,__)
```

(130) Nested Noun Phrases Analyzed in Two Stages

```
n.todo-np           NP!>>NP<(NP=np!$.__!$,__)
n.NP-ignore         NP!>>NP<(NP=np$.(FRAG|PRN|RBR|UH|":"|","!$.__)
                    !$,__)
```

```
n.PRP-NP          NP!>>NP<(NP</^PRP/$.(NP|NX=np!$.__)!$,__)
n.POS-NP          NP!>>NP<(NP<(POS!$.__)$.(NP|NX=np!$.__)!$,__)
n.filler-NP       NP!>>NP<(NP<(INTJ|UH!$.__)$.(NP|NX=np!$.__)!$,__)
n.DT-NP           NP!>>NP<(NP<(DT|PDT|/^RB/|JJR!$.__!$,__)$.(NP|
                  NX=np!$.__)!$,__)
n.Noun-DT         NP!>>NP<(NP=np$.(DT!$.__)!$,__)
n.NP-RB           NP!>>NP<(NP=np$.(RB!$.__)!$,__)
n.NP-QP           NP!>>NP<(NP=np$.(QP!$.__)!$,__)
n.NP-UCP          NP!>>NP<(NP=np$.(UCP!$.__)!$,__)
n.post-mod-ADJP   NP!>>NP<(NP=np$.(ADJP!$.__)!$,__)
n.ignore-ADVP     NP!>>NP<(NP=np$.(ADVP!$.__)!$,__)
```

C.2.2 Labeling Instrument Role Fillers

(131) Bare Role Fillers

```
b.bare            NP!>>NP<(NN|NNS|CD<(__=dep)!$.__!$,__)
b.DT-only         NP!>>NP<(DT|PDT|/^PRP/|/^RB/|INTJ$.(NN|NNS|CD
                  <(__=dep)!$.__)!$,__)
b.IN-that         NP!>>NP<(IN<that$.(NN|NNS|CD<(__=dep)!$.__)!$,__)
b.wh-words        NP!>>NP<(WDT|WP$.(NN|NNS|CD<(__=dep)!$.__)!$,__)
b.other-one-mod   NP!>>NP<(!DT|CD|/^PRP/|PDT|QP|/^RB/|INTJ|IN|"$"|
                  ADVP|PRN|CC|","|WDT|WP|/^JJ/|/^NN/|NP|/^VB/|ADJP|
                  UCP$.(NN|NNS|CD<(__=dep)!$.__)!$,__)
b.PDT-DT          NP!>>NP<(PDT$.(DT<(__=dep)!$.__)!$,__)
b.NN-RB           NP!>>NP<(NN|NNS<(__=dep)$.(RB!$.__)!$,__)
b.DT-VBN          NP!>>NP<(DT|/^PRP/$.(VBN<(__=dep)!$.__)!$,__)
b.pdets-NN        NP!>>NP<(PDT$.(DT$.(NN<(__=dep)!$.__))!$,__)
```

(132) Modified Role Fillers

```
m.Noun-PRN        NP!>>NP<(NNP<(__=dep)$.(PRN!$.__)!$,__)
m.CD-only         NP!>>NP<(CD|QP$.(NN|NNS|CD<(__=dep)!$.__)!$,__)
m.JJ-only         NP!>>NP<(/^JJ/|/^NN/|NP|/^VB/|ADJP$.(NN|NNS|CD
                  <(__=dep)!$.__)!$,__)
m.UCP-only        NP!>>NP<(UCP$.(NN|NNS|CD<(__=dep)!$.__)!$,__)
m.NN-PP           NP!>>NP<(NN|NNS<(__=dep)$.(PP!$.__)!$,__)
m.NN-QP           NP!>>NP<(NN<(__=dep)$.(QP!$.__)!$,__)
m.mod-VBN         NP!>>NP<(CD|NN$.(VBN<(__=dep)!$.__)!$,__)
m.leaf-postmod-1  NP!>>NP<(NN|NNS|VB|EX|JJ<(__=dep)$.(SBAR!$.__)!$,__)
m.leaf-postmod-3  NP!>>NP<(NN|NNS|VBZ<(__=dep)$.(S!$.__)!$,__)
m.NN-PUNCT-S      NP!>>NP<(NN<(__=dep)$.(!CC|PP|NN|NNS|/,/|JJ|DT
                  $.(S|SBAR!$.__))!$,__)
m.ADJP-NN         NP!>>NP<(__$.(ADJP|JJR|JJS$.(NN|NNS<(__=dep)
```

```
                  !$.__))!$,__)
m.NNP-NN          NP!>>NP<(__$.(NNP$.(NN|NNS<(__=dep)!$.__))!$,__)
m.VBN-NN          NP!>>NP<(__$.(VBG|VBN.(NN|NNS<(__=dep)!$.__))!$,__)
m.JJ-CD           NP!>>NP<(__$.(JJ$.(CD<(__=dep)!$.__))!$,__)
m.JJ-NN           NP!>>NP<(__$.(JJ$.(NN|NNS<(__=dep)!$.__))!$,__)
m.JJ-VB           NP!>>NP<(__$.(JJ$.(VB|VBG<(__=dep)!$.__))!$,__)
m.JJ-JJ           NP!>>NP<(__$.(JJ$.(JJ<(__=dep)!$.__))!$,__)
m.NN              NP!>>NP<(__$.(NN|NNS<(__=dep)$.(NN|NNS!$.__))!$,__)
m.NN-QP           NP!>>NP<(__$.(NN|NNS<(__=dep)$.(QP!$.__))!$,__)
m.post-SBAR       NP!>>NP<(__$.(NN|NNS<(__=dep)$.(S|SBAR!$.__))!$,__)
m.post-IN         NP!>>NP<(__$.(NN|NNS<(__=dep)$.(IN!$.__))!$,__)
m.post-JJ         NP!>>NP<(__$.(NN|NNS<(__=dep)$.(JJ!$.__))!$,__)
m.post-CD         NP!>>NP<(__$.(NN|NNS$.(CD<(__=dep)!$.__))!$,__)
m.final-quotes    NP!>>NP<(__$.(NN|NNS<(__=dep)$.(!NN|NNS|NNP|NP|
                  RB|CD|IN|JJ|POS|PP|PRP|QP|S|SBAR!$.__))!$,__)
m.compound-NP     NP!>>NP<(NP<(/^NN/|JJ|JJS|PP|CD|QP!$.__)$.(NP|
                  NX=np!$.__)!$,__)
m.NP-PP           NP!>>NP<(NP=np$.(PP!$.__)!$,__)
m.NP-WHPP         NP!>>NP<(NP=np$.(WHPP!$.__)!$,__)
m.NP-post-mod-1   NP!>>NP<(NP=np$.(SBAR!$.__)!$,__)
m.NP-post-mod-2   NP!>>NP<(NP=np$.(VP!$.__)!$,__)
m.NP-post-mod-3   NP!>>NP<(NP=np$.(S!$.__)!$,__)
m.NP-comma-PP     NP!>>NP<(NP=np$.(/,/$.(PP!$.__))!$,__)
m.NP-comma-NP     NP!>>NP<(NP=np$.(/,/$.(NP|ADVP|ADJP!$.__))!$,__)
m.NP-comma-S      NP!>>NP<(NP=np$.(/,/$.(S|SBAR|VP!$.__))!$,__)
m.NP-PUNCT-S      NP!>>NP<(NP=np$.(!CC|PP|NN|NNS|/,/|JJ|DT$.(S|
                  SBAR!$.__))!$,__)
m.NP-NN-PP        NP!>>NP<(NP=np$.(NN|NNS$.(PP!$.__))!$,__)
m.NP-PP-star      NP!>>NP<(NP=np$.(PP$.(__!$.__))!$,__)
```

(133) Nested Noun Phrases Analyzed in Two Stages

```
n.todo-np         NP!>>NP<(NP=np!$.__!$,__)
n.PRP-NP          NP!>>NP<(NP</^PRP/$.(NP|NX=np!$.__)!$,__)
n.POS-NP          NP!>>NP<(NP<(POS!$.__)$.(NP|NX=np!$.__)!$,__)
n.filler-NP       NP!>>NP<(NP<(INTJ|UH!$.__)$.(NP|NX=np!$.__)!$,__)
n.DT-NP           NP!>>NP<(NP<(DT|PDT|/^RB/|JJR!$.__!$,__)$.(NP|
                  NX=np!$.__)!$,__)
n.Noun-DT         NP!>>NP<(NP=np$.(DT!$.__)!$,__)
n.NP-RB           NP!>>NP<(NP=np$.(RB!$.__)!$,__)
n.NP-QP           NP!>>NP<(NP=np$.(QP!$.__)!$,__)
n.NP-UCP          NP!>>NP<(NP=np$.(UCP!$.__)!$,__)
n.post-mod-ADJP   NP!>>NP<(NP=np$.(ADJP!$.__)!$,__)
```

n.ignore-ADVP	NP!>>NP<(NP=np$.(ADVP!$.__)!$,__)
n.NP-comma-misc	NP!>>NP<(NP=np$.(/,/$.(!NP|S|SBAR|VP|ADJP|ADVP|PP!$.__))!$,__)

The author intended for this page to be blank.

Bibliography

Abe, Naoki, and Hang Li. 1996. Learning word association norms using tree cut pair models. *Proceedings of the 10th international conference on machine learning*, Bari, 3–11. 71

Adelman, James S.; Gordon D. A. Brown; and José F. Quesada. 2006. Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science* 17.814–823. 69, 103, 106, 126

Altmann, Gerry T.M., and Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73.247–264. 44, 151

Arnold, Jennifer E. 1998. *Reference form and discourse patterns*. Stanford University dissertation. 42

Arnold, Jennifer E. 2001. The effects of thematic roles on pronoun use and frequency of reference. *Discourse Processes* 31.137–162. 152

Arts, Anja; Alfons Maes; Leo G. M. Noordman; and Carel Jansen. 2011. Overspecification in written instruction. *Linguistics* 49.555–574. 44, 151

Baayen, R. H. 2011. *languager: Data sets and functions with "analyzing linguistic data: A practical introduction to statistics"*. R package version 1.4. Online: http://CRAN.R-project.org/package=languageR. 103, 126

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. 126

Barr, Dale J.; Roger Levy; Christoph Scheepers; and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255–278. 103, 126

Bates, Douglas; Martin Maechler; and Ben Bolker. 2013. *lme4: Linear mixed-effects models using s4 classes*. R package version 0.999999-2. Online: http://CRAN.R-project.org/package=lme4. 103, 126

Biber, Douglas; S. Johansson; Geoffrey N. Leech; Susan Conrad; and E. Finegan. 1999. *The Longman grammar of spoken and written English*. London: Longman. 53, 54, 59, 64

Bienvenue, B.; Gail Mauner; and Jean-Pierre Koenig. 2007. Anticipating instrument arguments. *Paper presented at the 20th annual CUNY conference on human sentence processing*, La Jolla, CA. 117

Bresnan, Joan. 1982. Control and complementation. *The mental representation of grammatical relations*, ed. by Joan Bresnan, 292–390. Cambridge, MA: MIT Press. 19

Brown, Paula M., and Gary S. Dell. 1987. Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology* 19.441–472. 43, 121

Brown, Paula Maureen. 1985. *Adapting production to comprehension: The explicit mention of instruments*. University of Rochester dissertation. 43, 44, 121

Browne, Wayles. 1971. Verbs and unspecified np deletion. *Linguistic Inquiry* 2.259–260. 155

Brunson, Barbara Anne. 1992. *Thematic discontinuity*. University of Toronto dissertation. 28

Burnard, L. 1995. *Users reference guide for the British National Corpus*. Oxford: Oxford University Computing Services. 1, 51

Chafe, Wallace L. 1972. Discourse structure and human knowledge. *Language comprehension and the acquisition of knowledge*, ed. by John B. Carroll and Roy O. Freedle, 41–69. Washington, D.C.: Winston. 41

Charniak, Eugene. 1995. Parsing with context free grammars and word statistics. Tech. Rep. CS-95-28, Brown University, Providence, Rhode Island. 52, 60

Charniak, Eugene. 1997. Statistical parsing with a context-free grammar and word statistics. *Proceedings of the fourteenth national conference on artificial intelligence*, 598–603. Menlo Park: AAAI Press/MIT Press. 52, 60

Ciaramita, Massimiliano, and Mark Johnson. 2000. Explaining away ambiguity: Learning verb selectional preference with bayesian networks. *Proceedings of the 18th international conference on computational linguistics*, Saarbrücken, 187–193. 71

Clark, Stephen, and David Weir. 2001. Class-based probability estimation using a semantic hierarchy. *Proceedings of the 2nd annual meeting of the north american chapter of the association for computational linguistics*, Pittsburgh, PA, 95–102. 71

Cohen, J. 1988. *Statistical power analyses for the behavioral sciences*. Hillsdale, NJ: Erlbaum. 78, 92, 93, 94, 97, 99, 119, 122, 124

Collins, Allan M., and M. Ross Quillian. 1969. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior* 8.240–247. 39

Cote, Sharon. 1992. Discourse functions of two types of null objects in English. *Presentation at the 66th annual meeting of the Linguistic Society of America*, Philadelphia, PA. 24

Dale, Robert, and Ehud Reiter. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science* 18.233–263. 42

de Marneffe, Marie-Catherine; Bill MacCartney; and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. *Lrec*. 150

Deerwester, S.; S. T. Dumais; G. W. Furnas; T. K. Landauer; and R. Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society of Information Science* 41.391–407. 139, 143

Dowty, David. 2003. The dual analysis of adjuncts/complements in categorial grammar. *Modifying adjuncts*, ed. by E. Lang, C. Maienborn, and C. Fabricius-Hansen, 33–66. Berlin: Mouton de Gruyter. 19

Elman, Jeffrey L. 2009. On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science* 33.1–36. 39, 143, 152

Engelhardt, Paul E.; Karl G. D. Bailey; and Fernanda Ferreira. 2006. Do speakers and listeners observe the Gricean maxim of quantity? *Journal of Memory and Language* 54.554–573. 148

Engelhardt, Paul E., and Fernanda Ferreira. 2012. Are listener's sensitive to articulation differences in over-described referring expressions: A test of the audience design hypothesis. *Poster presented at the 17th annual meeting of the architecture and mechanisms of language processing conference*, Paris, France. 43

Erk, Katrin; Sebastian Padó; and Ulrike Padó. 2010. A flexible, corpus-driven model of regular and inverse selectional preferences. *Computational Linguistics* 36.723–763. 143

Fellbaum, Christiane (ed.) 1998. *Wordnet: An electronic lexical database.* Bradford Books. 34, 71, 167

Fellbaum, Christiane, and Judy Kegl. 1989. Taxonomic structures and cross-category linking in the lexicon. *Proceedings of escol '89*, 93–104. 89

Ferreira, Victor S.; L. Robert Slevc; and Erin S. Rogers. 2005. How do speakers avoid ambiguous linguistic expressions? *Cognition* 96.263–284. 57, 58

Fillmore, Charles. 1968. The case for case. *Universals in linguistic theory*, ed. by Emmon Bach and Robert Harms, 1–87. New York: Holt, Rinehart and Winston. 28

Fillmore, Charles J. 1986. Pragmatically controlled zero anaphora. *Proceedings of the XII annual meeting of the berkeley linguistics society, bls*, ed. by V. Nikiforidou, M. Vanllay, M. Niepokuj, and D. Felder, Berkeley, CA. 24, 89, 155

Firth, John R. 1968. *Selected papers of J. R. Firth 1952–59*. Bloomington: Indiana University Press. F. R. Palmer (ed.). 14, 45, 136

Foley, William, and Robert Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press. 19

Francis, W., and H. Kučera. 1982. *Frequency analysis of English usage*. New York: Houghton Mifflin Co. 26, 91

Gildea, Daniel, and Daniel Jurafsky. 2000. Automatic labeling of semantic roles. *Processing of acl-00*, Hong Kong, 512–520. 150

Gildea, Daniel, and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics* 28.245–288. 150

Grice, Paul. 1975/2002. Logic and conversation. *Foundations of cognitive psychology: Core readings*, ed. by Daniel J. Levitin, 719–732. MIT Press. 15, 44

Grodner, D., and J. Sedivy. 2011. The effects of speaker-specific information on pragmatic inferences. *The processing and acquisition of reference*, ed. by N. Pearlmutter and E. Gibson. Cambridge, MA: MIT Press. 44, 151

Grosz, B. J.; A. K. Joshi; and S. Weinstein. 1995. Centering: a framework for modelling the local coherence of discourse. *Computational Linguistics* 21.203–226. 41

Grosz, Barbara J.; Aravind K. Joshi; and Scott Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. *Proceedings of the 21st annual meeting on association for computational linguistics*, ACL '83, 44–50. Stroudsburg, PA, USA: Association for Computational Linguistics. 41

Grosz, Barbara J., and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12.175–204. 41

Gundel, Jeanette K.; Nancy Hedberg; and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69.274–307. xi, 41

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. *The second meeting of the north american chapter of the asssociation for computational linguistics*. 145

HALE, JOHN. 2006. Uncertainty about the rest of the sentence. *Cognitive Science* 30.643–672. 145

HANKAMER, JORGE. 1973. Unacceptable ambiguity. *Linguistic Inquiry* 4.17–68. 26

HEBB, DONALD O. 1949. *The organization of behavior.* New York: Wiley & Sons. 39

VAN HEUVEN, WALTER J. B.; PAWEL MANDERA; EMMANUEL KEULEERS; and MARC BRYSBAERT. 2013. Subtlex-uk: A new and improved word frequency database for british english. *The Quarterly Journal of Experimental Psychology* 0.1–36. 69, 70

HOFMEISTER, PHILIP. 2010. Semantic processing and memory retrieval. *Presentation at the linguistic society of american annual meeting*, Baltimore, MD. 147

HOPPER, PAUL, and SANDRA THOMPSON. 1980. Transitivity in grammar and discourse. *Language* 56.251–299. 155

HOROWITZ, ALEXANDRA, and MICHAEL C. FRANK. 2012. Learning from speaker word choice by assuming adjectives are informative. *Proceedings of the 34th annual meeting of the cognitive science society.* 44, 151

JACKENDOFF, RAY. 1987. The status of thematic relations in linguistic theory. *Linguistic Inquiry* 18.369–411. 28

JAEGER, T. F. 2005. Optional that indicates production difficulty: Evidence from disfluencies. *Proceedings of DiSS'05, disfluency in spontaneous speech workshop*, Aix-en-Provence, France, 103–109. 146

JONES, MICHAEL N., and DOUGLAS J. K. MEWHORT. 2007. Representing word meaning and order information in a composite holographic lexicon. *Psychological Review* 114.1–37. 139

JURAFSKY, DANIEL. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20.137–194. 145

KEENAN, EDWARD, and BERNARD COMRIE. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8.53–60. 141

KINTSCH, WALTER. 2001. Predication. *Cognitive Science* 25.173–202. 139

KOENIG, JEAN-PIERRE; GAIL MAUNER; and BRETON BIENVENUE. 2003. Arguments for adjuncts. *Cognition* 89.67–103. 11, 19, 22, 23, 30, 80, 110, 112, 117

KOENIG, JEAN-PIERRE; GAIL MAUNER; BRETON BIENVENUE; and KATHY CONKLIN. 2008. What with? the anatomy of a (proto)-role. *Journal of Semantics* 25.175–220. 28, 29, 30, 110, 112, 118, 138

KOOLEN, RUUD; ALBERT GATT; MARTIJN GOUDBEEK; and EMIEL KRAHMER. 2011. Factors causing overspecification in definite descriptions. *Journal of Pragmatics* 43.3231–3250. 122

LAKOFF, GEORGE. 1968. Instrumental adverbs and the concept of deep structure. *Foundations of Language* 4.4–29. 28

LEE, LILLIAN. 1997. *Similarity-based approaches to natural language processing*. Harvard University dissertation. 143

LEHRER, ADRIENNE. 1970. Verbs and deletable objects. *Lingua* 25.227–254. 24, 25, 82, 89

LEVELT, WILLEM J.M. 1989. *Speaking: From intention to articulation*. Cambridge/London: MIT Press. 43

LEVIN, BETH. 1993. *English verb classes and alternations*. Chicago: Chicago University Press. 21, 24, 25, 82

LEVY, ROGER. 2008. Expectation-based syntactic comprehension. *Cognition* 106.1126–1177. 145

LEVY, ROGER, and T. FLORIAN JAEGER. 2007. Speakers optimize information density through syntactic reduction. *Proceedings of the twentieth annual conference on neural information processing systems*. 145

LIN, DEKANG. 1997. Using syntactic dependency as local context to resolve word sense ambiguity. *Proceedings of acl-97*, Madrid, Spain. 139

LIN, DEKANG. 1998a. Automatic retrieval and clustering of similar words. *Proceedings of the joint annual meeting of the association for computational linguistics and international conference on computational linguistics*, Montréal, Canada, 768–774. 139

LIN, DEKANG. 1998b. Dependency-based evaluation of MINIPAR. *Proceedings of the LREC workshop on the evaluation of parsing systems*, Granada, Spain, 234–241. 150

LUND, K.; C. BURGESS; and R. A. ATCHLEY. 1995. Semantic and associative priming in high-dimensional semantic space. *Cognitive science proceedings, lea*, 660–665. 139

MACDONALD, M.C.; J. PEARLMUTTER; and MARK S. SEIDENBERG. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101.676–703. 19

MANNING, C. 2003. Probabilistic syntax. *Probabilistic linguistics*, ed. by R. Bod, J. Hay, and S. Jannedy, 289–314. Cambridge, MA: MIT Press. 19

MATSUKI, KAZUNAGA; TRACY CHOW; MARY HARE; JEFFREY L. ELMAN; CHRISTOPH SCHEEPERS; and KEN MCRAE. 2011. Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37.913–934. 152

McCawley, James. 1968. The role of semantics in a grammar. *Universals in linguistic theory*, ed. by Emmon Bach and Robert Harms, 124–169. Holt: Rinehart and Winston. 26, 35

Miller, George A. 1995. WordNet: A lexical database for English. *Communications of the ACM* 38.39–41. 71

Mittwoch, A. 1971. Idioms and unspecified np deletion. *Linguistic Inquiry* 2.255–259. 155

Mittwoch, A. 1982. On the difference between *eating* and *eating something*: activities versus accomplishments. *Linguistic Inquiry* 13.113–122. 155

Newman, Matthew L.; Carla J. Groom; Lori D. Handelman; and James W. Pennebaker. 2008. Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes* 45.211–236. 78

Nilsen, Don L. F. 1973. *The instrumental case in English: Syntactic and semantic considerations*. The Hague: Mouton. 28

Padó, Sebastian, and Mirella Lapata. 2003. Constructing semantic space models from parsed corpora. *Proceedings of the 41st annual meeting of the association for computational linguistics, acl'03*, 128–135. 139, 140

Padó, Sebastian, and Mirella Lapata. 2007. Dependency-based construction of semantic space models. *Computational Linguistics* 33.161–199. 141

Pennebaker, J. W.; M. E. Francis; and R. J. Booth. 2001. *Linguistic inquiry and word count (liwc): Liwc2001*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc. 78

Pollard, Carl, and Ivan Sag. 1987. *Information-based syntax and semantics*, vol. 1. Stanford: CSLI Publications. 19

R Core Team. 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Online: `http://www.R-project.org/`. 103, 126

Rappaport Hovav, Malka, and Beth Levin. 2001. An event structure account of English resultatives. *Language* 77.766–797. 157

Rappaport Hovav, Malka, and Beth Levin. 2010. Reflections on manner/result complementarity. *Syntax, lexical semantics, and event structure*, ed. by E. Doron, Malka Rappaport Hovav, and I. Sichel, 21–38. Oxford, UK: Oxford University Press. 157

Resnik, Philip. 1996. Selectional constraints: An information-theoretic model and its computational realization. *Cognition* 61.127–159. 71

Resnik, Philip Stuart. 1993. *Selection and information: A class-based approach to lexical relationships*. University of Pennsylvania dissertation. xi, 10, 21, 24, 26, 31, 33, 37, 46, 71, 80, 82, 83, 89, 90, 91, 93, 95, 110, 111, 116, 120, 137, 141, 170

Rice, Sally. 1988. Unlikely lexical entries. *Proceedings of the berkeley linguistics society*, 202–212. 89

Rissman, Lilia. 2010. Instrumental with, locatum with and the argument/adjunct distinction. *LSA Annual Meeting Extended Abstracts*. Online: `http://elanguage.net/journals/index.php/lsameeting/article/view/629/613`. 19

Rissman, Lilia. 2013. *Event participant representations and the instrumental role: a cross-linguistic study*. Johns Hopkins University dissertation. 19

Rohde, Hannah. 2008. *Coherence-driven effects in sentence and discourse processing*. University of California, San Diego dissertation. 152

Rohde, Hannah; Andrew Kehler; and Jeff Elman. 2006. Event structure and discourse coherence biases in pronoun interpretation. *Proceedings of the 28th annual conference of the cognitive science society*, Vancouver. 152

Rohde, Hannah; Andrew Kehler; and Jeff Elman. 2007. Pronoun interpretation as a side effect of discourse coherence. *Proceedings of the 29th annual conference of the cognitive science society*, Nashville, TN, 617–622. 152, 153

Roland, Doug; Frederic Dick; and Jeffrey L. Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57.348–379. 60, 175

Roland, Douglas; Hongoak Yun; Jean-Pierre Koenig; and Gail Mauner. 2012. Semantic similarity, predictability, and models of sentence processing. *Cognition* 122.267–279. 143

Roland, Douglas William. 2001. *Verb sense and verb subcategorization probabilities*. University of Colorado dissertation. 150

Sag, Ivan. 1976. *Deletion and logical form*. MIT dissertation. 26

Sahlgren, Magnus. 2006. *The word-space model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. Stockholm University dissertation. Department of Linguistics. 143

Sedivy, J. C. 2006. Evaluating explanations for referential context effects: Evidence for Gricean mechanisms in online language interpretation. *In approaches to world-situated language use: Bridging the language-as-product and language-as-action traditions (learning, development, and conceptual change*, ed. by J. C. Trueswell and M. K. Tanenhaus. Cambridge, MA: MIT Press. 44, 151

Sedivy, Julie C. 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research* 32.3–23. 44, 151

Sedivy, Julie C.; Michael K. Tanenhaus; Craig G. Chambers; and Gregory N. Carlson. 1999. Achieving incremental processing through contextual representation: Evidence from the processing of adjectives. *Cognition* 71.109–147. 44, 151

Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27.379–423, 623–656. 34

Simmons, Robert F. 1973. Semantic networks: Their computation and use for understanding English sentences. *Computer models of thought and language*, ed. by Roger C. Schank and Kenneth Mark Colby, 61–113. San Francisco: W.H. Freeman and Co. 150

Sinclair, John (ed.) 1987. *Collins COBUILD English language dictionary*. London: Collins. 90

Spivey, Michael J., and Daniel C. Richardson. 2008. Language embedded in the environment. *The cambridge handbook of situated cognition*, ed. by P. Robbins and M. Aydede, 383–400. Cambridge, UK: Cambridge University Press. 43

Stevenson, Rosemary J.; Rosalind A. Crawley; and David Kleinman. 1994. Thematic roles, focus and the representation of events. *Language and Cognitive Processes* 9.473–592. 152

Strzalkowski, Tomek (ed.) 1999. *Natural language information retrieval*. Dordrecht: Kluwer Academic Publishers. 139

Tesnière, L. 1959. *Elèments de syntaxe structurale*. Paris: Klincksieck. 19

Turney, P. 2001. Mining the web for synonyms: PMI-IR versus LSA on TOEFL. *Proceedings of the twelfth european conference on machine learning (ecml-2001)*, ed. by L. De Raedt and P. Flach, Freiburg, Germany, 491–502. 149

van der Sluis, Ielka, and Emiel Krahmer. 2007. Generating multimodal references. *Discourse Processes* 44.145–174. 42, 43, 97, 98, 100, 107, 110

Vendler, Z. 1967. *Linguistics in philosophy*. Ithaca: Cornell University Press. 155

Wiemer-Hastings, P. 2000. Adding syntactic information to lsa. *Proceedings of the 22nd annual conference of the cognitive science society*, 989–993. Mahwah, NJ: Erlbaum. 139

Williams, Edwin. 1977. Discourse and logical form. *Linguistic Inquiry*, 101–139. 26

Wurm, Lee H., and Sebastiano A. Fisicaro. 2014. What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language* 72.37–48. 101, 126

Yun, Hongoak. 2012. *Predictability versus activation of representations: Evidence from sentence comprehension.* University at Buffalo, State University of New York dissertation. 30, 112, 120

Yun, Hongoak; Gail Mauner; and Jean-Pierre Koenig. 2006. Anticipation vs. integration of syntactically infrequent but semantically obligatory arguments. *Poster presented at the 19th annual CUNY conference on human sentence processing*, CUNY, NY. 80, 117