Take a Look at This!

Form, Function and Productivity of English Light Verb Constructions

by

CLAIRE NICOLE BONIAL

B.A., New Mexico State University, 2004

M.A., University of Colorado Boulder, 2007

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Linguistics

2014

This thesis entitled:
Take a Look at This! Form, Function and Productivity of English Light Verb Constructions
written by Claire Nicole Bonial
has been approved for the Department of Linguistics

_____

Martha Palmer

_____

Bhuvana Narasimhan

_____

Albert Kim

_____

Laura Michaelis

_____

Suzanne Stevenson

Date: _____

The final copy of this thesis has been examined by the signatories, and we
find that both the content and the form meet acceptable presentation standards
of scholarly work in the above mentioned discipline.

Bonial, Claire Nicole (Ph.D., Linguistics, Department of Linguistics)

Take a Look at This! Form, Function and Productivity of English Light Verb Constructions

Thesis directed by Professor Martha Palmer & Associate Professor Bhuvana Narasimhan

English light verb constructions (LVCs), such as *have a drink, make an offer, take a bath, do an investigation*, and *give a groan*, represent a powerfully expressive resource of English; however, the definition, linguistic function and productivity of English LVCs remain unclear. This research focuses on exploring these three issues. A definition for LVCs that combines syntactic and semantic criteria is given after a survey of existing research on delimiting and defining LVCs. This definition is implemented in the development of a LVC annotation schema for the PropBank project, and these annotations are in turn used in the development of the state-of-the-art automatic system for identifying LVCs. Existing research on the linguistic function of LVCs both cross-linguistically and in English is analyzed, and a corpus study provides evidence that the primary function of LVCs in English is to enable speakers to describe events in a manner that can take advantage of rich nominal modification, for example, *The inspector general did a **rather controversial** investigation...* Finally, linguistic and cognitive approaches to the development of grammar and the extension of constructions are discussed, and the hypothesis that novel constructions are extended by semantic analogy to an existing, highly frequent exemplar is tested in the domain of LVCs, using large-scale Mechanical Turk surveys. In closing, the potential impact of these findings on both Natural Language Processing and linguistic theory are presented, as well as opportunities for future work.

Dedication

For my parents, Ron and Mary Bonial,

who taught me to love and respect the power of words,

whether Shakespeare or puns.

Acknowledgments

This is very much a collaborative work, and I owe thanks to so many. I would firstly like to extend my gratitude to my advisors, Martha Palmer and Bhuvana Narasimhan. Without Martha's bedrock of research, and her unflagging support and encouragement throughout my time at CU, this work would not have been possible. Thank you to Bhuvana for sharing her incredible ability to understand and express the nuanced variables involved in every linguistic problem. You have both truly been mentors to me, and I hope to carry on your tradition of excellence in both research and teaching. I also owe much gratitude to Suzanne Stevenson, who helped me grow the initial seed of this work into dissertation research and beyond. Thank you to Laura Michaelis for her invaluable input in the areas of syntax, aspect and construction grammar. I would also like to thank Al Kim for his expert advice in the cognitive aspects of this work.

I have had such wonderful opportunities for interdisciplinary collaboration in my time at CU, and I would like to extend my deepest gratitude for the opportunity to collaborate with Wei-Te Chen in the development of an automatic classifier of English light verb constructions. I am also very grateful to the assistance and support from fellow students Julia Bonn, Alexis Raykhel, and Katie Conger, who have all made my work more fun than work! I would also like to thank Afsaneh Fazly, for her generous assistance in the early stages of this project, and for bringing her own expertise in the area of light verb constructions to my research.

Finally, thanks to all of my friends and family, who have kept my life balanced, and me smiling and laughing even in the most stressful times. I give you all my thanks!

Table of Contents

Tables

Table

Figures

Figure

Charts

Charts

Chapter 1
Introduction & Background

1.1 Introduction

Consider for a moment the simple, everyday English phrases, *Sarah took a bath* and *Sarah gave the baby a bath.* What events are denoted by these phrases? If you were not a proficient English speaker, you might look up the words in the sentence. The first definition for *take* appearing in the American Heritage Dictionary[1] is "to get into one's possession..." The first definition for *give* is "to make a present of." Thus, one could logically conclude that *giving* and *taking* events involve the transfer of possessions. What object is being transferred here? A *bath*: "a vessel for holding water in which to wash the body." Yet, for fluent speakers of English, these phrases do not conjure an image wherein Sarah is picking up a bath tub and either taking it somewhere or giving it to a baby; rather, both phrases refer to events where Sarah either bathed herself or bathed the baby.

These phrases are often referred to as "Light Verb Constructions" (LVCs) (Jespersen, 1942); however, both the labeling and definition of such constructions remain under debate in the linguistic community. English LVCs also include expressions like *have a drink, make an offer, take a walk,* and *do an investigation,* and are usually thought to consist of a semantically general verb and a noun that denotes an event or state. Such expressions can be problematic for Natural Language Processing (NLP) because the lexical resources used by many NLP applications provide static representations for verb semantics, but LVCs are one manner in which speakers can shift and extend verb semantics in novel ways. For example, VerbNet (Kipper et al., 2008)

---

[1] http://www.ahdictionary.com

provides information on the syntax and semantics of both *take* and *give*.[2] While the coverage of

take senses is quite extensive, and *give* is appropriately represented in a class of transfer verbs,

the "light" senses of the verbs are not represented in VerbNet. In part, this is because the event

semantics of such constructions stem primarily from the noun portion of the expression, rather

than the verb. To address this, another lexical resource, FrameNet (Baker et al., 1998; Fillmore,

Johnson & Petruck, 2002), handles this fact by listing the light verb with the noun in their

annotations of noun relations. Similarly, WordNet (Fellbaum, 1998) gives example sentences for

noun entries that may or may not include accompanying light verbs. Nonetheless, it is difficult if

not impossible for such resources to keep up with LVCs by simply listing them because speakers

will continuously introduce novel LVCs. Thus, it is necessary for us to understand not only how

to recognize an LVC and its unique semantics, but also to begin to understand how we can

predict what novel *light verb + noun* combinations could arise to create grammatical LVCs.

1.1.1 What are LVCs?

Definitions of LVCs both cross-linguistically and within a language remain nebulous, and this

has in turn hampered efforts to understand the function and semantics of LVCs. Since

Jespersen's (1942, Volume VI: 117) application of the term "light verb" to English *verb + noun*

phrase constructions like *have a rest*, the term has been extended to constructions with Japanese

suru 'do' (Grimshaw & Mester, 1988), Romance causatives (Rosen, 1989), Hindi *noun + verb*

constructions (Mohanan, 1994), Urdu *verb + verb* constructions (Butt, 1995), as well as a

Chinese variant on control/raising constructions involving *ba* and *de* (Huang, 1992). While these

---

[2] https://verbs.colorado.edu/verb-index/

varying linguistic phenomena share a common core of two or more predicational elements, little

unites the definition beyond that. As Butt (2003) points out, LVCs can only be distinguished in a

language-dependent fashion. Thus, while in some languages there is a clear syntactic definition

of LVCs (e.g. Japanese, Urdu), in other languages one must rely more heavily on a semantic

definition (e.g. English). Underlying this is the nagging issue of whether or not one should

consider these somewhat varying cross-linguistic phenomena under the same definition at all.

Furthermore, some theories do not distinguish LVCs as a separate type of construction at all, but

rather as a type of auxiliary (Hopper & Traugott, 1993), or simply another idiom (Givón, 1979).

In this research, the constructions of focus consist of a semantically general, highly

polysemous verb and a noun that denotes an event or state. These two elements together create a

complex predicate: while the noun provides the majority of the event semantics, the verb both

syntactically licenses this, and modulates the meaning of the expression in a variety of manners

and extents depending upon the exact LVC. The verbs allowed in these constructions have

relatively low "cue validity" in the sense expressed by Goldberg (2006:109), meaning that the

verb alone is not a very reliable predictor of overall sentence meaning. Nonetheless, the

contribution of the verb in the overall meaning of the complex predicate is easily seen when we

compare two LVCs with the same noun but a different verb; for example:

1. Sarah **gave** the baby a bath. (Sarah bathed the baby)
2. Sarah **took** a bath. (Sarah bathed herself)

Here, the choice of light verb can alter the valency of the complex predicate. However, it is the noun that provides most of the information about the event or state (i.e. expresses the predication of the utterance).

These constructions are distinct from "heavy" usages of the same verb. A "heavy" usage of the verb is one in which the verb sense is a canonical, literal sense. For example, the heavy sense of *make* is *create*, *take* is *cause motion/get into possession, have* is *possession* and *give* is *transfer possession* (e.g. *make a cake, take a book off the shelf, have a bicycle, give a present*). In their heavy usages, the sentence can generally be interpreted compositionally: each component word contributes its lexical semantics to the sentence meaning. LVCs*, such as *have a drink* and *take a walk*, are neither purely idiomatic nor purely compositional in their meaning. They are somewhat idiomatic because their interpretation relies upon recognizing that the semantics of these phrases, including the roles of participants in the events, stem largely from the nominal complement rather than the verb. For example, the aforementioned phrases, *have a drink* and *take a walk* do not primarily express events of possession or transfer of possession, as one may expect when considering the verbs *have* and *take* respectively. Rather, these phrases express events of *drinking* and *walking.* Thus, despite the fact that these phrases are syntactically indistinguishable from phrases like *have a toy* or *take a toy,* these particular combinations of words carry special meaning. In this way, LVCs are similar to idioms like *trip the light fantastic* ('dance'). However, LVCs clearly differ from fixed, non-compositional idioms like *kick the bucket,* in which the meaning of the phrase ('die') has no clear relationship to the component words of the phrase. The meaning of the nominal complement within an LVC is quite transparent; it is the combined meaning of the verb and noun that is idiomatic, and arguably

changeable depending upon the verb and noun combined. Thus, LVCs are neither fully idiomatic nor fully compositional, and can best be thought of as semi-compositional.

1.1.2 Research Motivation: Why are (English) LVCs important?

English LVCs in particular pose interesting theoretical questions because cross linguistically they are somewhat unusual: it is generally possible to define LVCs syntactically according to language-dependent criteria (Butt, 2003, 1995; Grimshaw & Mester, 1988), and it is common for LVCs to exist in complementary distribution with lexical verbs, such that LVCs are exploited where a lexical verb counterpart for the concept doesn't exist (Zarco, 1999). For example, drawing on the work of Schutze-Berndt (2000), Butt (2003) points out that light verbs in the Australian language Jaminjung can be recognized based on their distribution. Jaminjung has a class of coverbs, which do not inflect and seem to share characteristics with both verbs and adverbs. Additionally, the language has a closed class of inflecting verbs that can be used to predicate as a main verb, but when combined with a coverb, their predicational power is light (below, "go" is the inflecting verb with light semantics, while "race" is the coverb):

3.	burdurdubba=biya	ga-ngga	ngayin	thanthu
	race=now		3Sg-go.Pres	animal(Abs)	Dem(Abs)
	'It is racing off now that animal.'
	(Jaminjung, Schultze-Berndt 2002; cited in Butt, 2003:6)

Thus, the light verb usages can be recognized by their patterning with coverbs.

Northern Australian languages like Wagiman (Wilson, 1999), Bardi (Bowern, 2002), and Jaminjung bring to light an interesting feature of LVCs crosslinguistically. These languages make extensive use of LVCs, and this seems to be related to the fact that there is a relatively small, closed class of inflecting predicates. Thus, LVCs fill a void in the language's resources, extending the array of meanings that can be expressed with a closed set of inflecting predicates. In contrast, English LVCs aren't syntactically distinguishable from other *verb + noun* constructions, and have lexical verb counterparts that are semantically very similar (e.g. *Sarah made an offer to buy the house, Sarah offered to buy the house*), which raises the question of why the two variants continue to exist together in the language.

Furthermore, recognizing and understanding LVCs has both theoretical importance in linguistics generally, and practical importance in NLP. LVCs are theoretically interesting and challenging to NLP partly because the constructions are also "semi-productive" (Nickel, 1978: 83), meaning that speakers can use the pattern of *light verb + eventive/stative noun* productively to create novel LVCs. However, this productivity is constrained because not all combinations are thought to be acceptable. As semi-productive constructions, LVCs share some characteristics with idioms (most of which are not productive) and share other characteristics with purely collocational language (which is theoretically fully productive). For example, as previously mentioned, LVCs are similar to idioms in that their meanings are not entirely predictable from their component parts. However, unlike a fixed idiom like *kick the bucket,* LVCs do have some measure of meaning that stems from the lexical semantics of its component parts, and there is a limited range of flexibility in the substitutability of terms within the LVC. There are no other possible terms that can be substituted in a fixed idiom to maintain the semantics: one cannot

6

choose to say, "He struck the pail" instead of "He kicked the bucket." In the case of LVCs, some

substitution is possible, as evidenced by families of semantically similar LVCs:

4.   Sarah made a proclamation/speech/announcement.
5.   Sarah made a recommendation/suggestion to me.

However, the construction is not *fully* productive because other similar combinations do not

create LVCs that are considered acceptable to native speakers; they simply sound "funny":

6.   ?Sarah made a yell.[3]
7.   ?Sarah made advice to me.

As semi-productive constructions, LVCs present unique and important evidence for the

development and organization of grammar, and specifically for considering views of

Construction Grammar (see for example, Goldberg, 1996, 2006; Croft & Cruse, 2004) and other

usage-based theories, such as Emergent Grammar (see, for example, Hopper, 1998; Bybee and

Hopper, 2001; MacWhinney, 2001; Bybee, 2010; Bybee and McClelland, 2005).

In the domain of NLP, the successes of supervised approaches to automatic semantic

analysis (e.g. Màrquez et al., 2008) rely on quality lexical resources such as PropBank, VerbNet,

and FrameNet, which provide information on the relational semantics of verbs. However,

automatic systems are limited by the static nature of these resources, which do not currently

reflect the potential for speakers to use a verb in new contexts. As semi-productive

---

[3] Throughout this work, phrases preceded by *?* are questionably grammatical, while phrases preceded by * are likely ungrammatical.

constructions, LVCs are one example of a construction in which a verb's semantics can be extended and shifted as it combines with another element (here, with a noun predicate). To supplement lexical resources with LVCs, it is necessary to define and understand the function of LVCs as an alternative to lexical verbs. Additionally, to make these resources more dynamic in their treatment of LVCs and other semi-productive constructions, we must better understand how such constructions are extended to novel combinations.

1.2 Research Questions & Dissertation Structure

This research addresses several questions concerning the nature of LVCs. Firstly, an attempt is made to answer the question of how to best define LVCs and therefore delimit them from other constructions. In order to address this question, existing definitions are compared and critiqued, culminating in a definition of LVCs that is both syntactic and semantic. This definition is currently in use in the PropBank (Palmer et al., 2005) annotation project to identify LVCs, and these annotations have been used to develop an automatic classifier of LVCs, which will be described and compared to other existing classifiers. Secondly, the question of the function of LVCs, especially in contrast to counterpart lexical verbs (e.g. *take a walk, walk*), is investigated. The approach to this question firstly involves a survey of existing research on this question, and a corpus study of the behavior and distribution of PropBank LVCs in comparison to their lexical verb counterparts. This study demonstrates that while LVCs generally expand the expressive power of English, the primary function of LVCs is to allow for rich event description via nominal modification. The final question addressed by this research concerns the nature of LVC

productivity. LVCs seem to have identifiable "families" or clusters of semantically similar LVCs

(e.g. *make a speech, make a declaration, make proclamation*); thus, the hypothesis that novel

LVCs are semantically similar extensions from a highly frequent exemplar is tested through

acceptability judgments gathered on Mechanical Turk.

In summary, the research questions addressed here are as follows:

1. **Defining LVCs:** How do we best define LVCs, and therefore delimit LVCs from other
   constructions?
2. **LVC function:** What is the function of LVCs, especially in contrast to their lexical verb
   counterparts?
3. **LVC productivity:** What is the nature of the productivity of LVCs? Specifically, how do
   novel LVCs enter the language, and why do there appear to be "families" of semantically
   similar LVCs?

These three research questions are addressed in the order given above, and will be examined with

respect to related work in computational linguistics, psycholinguistics, and cognitive science.

The structure of the dissertation follows. In the remainder of this introductory chapter

(Section 1.3), an overview of central theories and assumptions is given. In Chapter 2, the

question of how to define and delimit LVCs is analyzed firstly through existing work in

linguistics and Natural Language Processing, then the PropBank guidelines for detection and

annotation are described, as well as the automatic system for LVC detection. In Chapter 3, the

function of English LVCs is examined through the lens of synchronic, semantics-based theories,

compared with views from diachronic research on the topic. These views are synthesized

through a corpus study of LVCs, which compares several key features of the constructions to

their counterpart lexical verbs. In Chapter 4, the productivity of LVCs is firstly situated in

theories of the development of grammar generally, and then in theories of the development and

extension of semi-productive constructions. A dominant hypothesis that novel constructions are

formed by semantic analogy to an existing, highly frequent exemplar is introduced, and

acceptability judgments testing this hypothesis are described. Chapter 5 concludes the

dissertation with a discussion of the broader impacts of this research; opportunities for future

work are also presented there.

1.3 Theoretical Background

This dissertation is largely descriptive and does not attempt a formal syntactic model of LVCs.

As such, a construction-based approach to this research was found to have several advantages.

Firstly, a construction-based approach does not require a single licensing head. Instead, two

elements can co-license a particular structure. For example, in noun phrases, it is not clear what

the head of the phrase is: does the noun license a particular determiner, or does the determiner

license a particular type of noun (mass/count)? In a construction-based account, one can think of

the noun and article as co-licensing (Michaelis, 2004). In the case of LVCs, a classic phrase

structure account would predict that the verb is the head, and licenses arguments that are the

sisters to the verb and verb phrase. However, it seems quite clear from the meaning of the

constructions that the semantic roles assigned in LVCs are assigned by the noun, not the verb.

Formal syntactic theories have accounted for this by positing processes of argument transfer

from the noun to the verb's arguments, or by exploring the possibility of two heads in the case of

LVCs (Grimshaw & Mester, 1988; Cattell, 1984). While these accounts provide evidence for why only certain verbs can be used as light verbs (verbs that are hypothesized to have incomplete argument structures), these accounts do not attempt an explanation of the semi-productivity of LVCs. Thus, even if one posits a theory of argument transfer or dual heads, these formal syntactic accounts fail to pinpoint why certain combinations of *light verb + noun* are acceptable, while others are not (e.g. *make a declaration, ?make a yell)*. In a construction-based approach, the verb and noun can be understood to co-license the construction; thus, each can bring distinct constraints on their combination that may result in the idiosyncratic patterns of LVC productivity.

Secondly, and relating to the last point regarding idiosyncrasy, a construction-based model allows us to view the grammar as an idiomaticity continuum, using an array of constructions of correspondingly graded generality. As a result, where other approaches may focus on one extreme of the continuum (for example, focusing on open schemas of the type easily modeled by phrase structure rules in a rule-to-rule approach), the construction-grammar approach embraces an account of structures that acknowledges the fuzzy boundaries between idiomatic patterns and general (transparent, lexically unspecified) patterns (Kay & Michaelis, 2012). As semi-productive, semi-idiomatic constructions, LVCs are on this cline; thus, it is helpful to view LVCs through the lens of a theory that represents such patterns with the same mechanisms that are used to model the open schemas.

Finally, a construction-based approach is a usage-based approach, espousing the idea that language structure arises from usage. Usage-based approaches devote attention to frequency of usage as well, which plays an important role in the development of grammar in these approaches.

As frequency has been used in explanations of how humans learn the idiosyncratic facts of

language (e.g. Goldberg, 2011), an approach acknowledging the role of frequency is also helpful

when examining LVCs. Although researchers in other traditions have attempted to account for

the unique semantics of idioms and other morphological and syntactic irregularities of language

(Fraser, 1970; Jackendoff, 1975; Nickel, 1978), they do not necessarily devote the same level of

empirical study to such phenomena as the usage-based approaches do. A brief introduction and

background to these approaches are given in the paragraphs to follow. Further details are

provided as appropriate in later chapters of the dissertation.

1.3.1 Usage-Based, Emergent Grammar

Emergent Grammar (see, for example, Hopper, 1998; Bybee and Hopper, 2001; MacWhinney,

2001; Bybee, 2010; Bybee and McClelland, 2005) emphasizes the functional nature of language:

syntactic facts are never independent of the function of a particular form or utterance. Hopper's

theory of Emergent Grammar (1998) challenges the assumption in traditional linguistics that

there is an abstract, mentally represented rule system that is implemented when we speak.

Instead, grammar (if it can still be called as much), like culture, should be viewed "as a real-time,

social phenomenon" (1998: 141). As such, grammar is always temporary and dynamically

emergent; a final, static grammar is never attained. The regularity that can be observed in

language stems from discourse and, in turn, constantly shapes discourse. Thus, while traditional

grammars often seem to assume that some knowledge of grammar is a prerequisite for discourse,

Emergent Grammar assumes that structures emerge from specific, concrete interactions. Hopper

explains,

"The notion of Emergent Grammar is meant to suggest that structure, or regularity, comes out of discourse and is shaped by discourse in an ongoing process. Grammar is, in this view, simply the name for certain categories of observed repetitions in discourse" (1998: 156).

The sort of fixed templates of recurring linguistic patterns that are observed are anchored in a speaker's experience with specific instances of that utterance. Everyday language is therefore built up of prefabricated parts of such templates that form a speaker's individual discourse experience. These templates can be thought of as "constructions," which are generally thought of as any pairing of form and meaning, including phrases.

Concerning productivity of certain constructions, Hopper proposes that "constructions spread outwards from a small nucleus and in turn form new nuclei," this process results in utterances that are in "family resemblance relationships to one another" (Hopper, 1998: 150). Constructions that are more useful in the language will attain a more consistent structure, and will more often serve as a basis for variation and extension. Similarly, Bybee (2010) proposes a process wherein novel constructions are extended by semantic analogy to existing, high-frequency constructions. This picture could fit well with some of the observed tendencies of LVC productivity. Presumably, an LVC arises in the discourse context when it is needed, either because an appropriate lexical verb doesn't exist in the language, or the speaker wants to convey shades of meaning that aren't easily expressed using verbal modification, and are more easily expressed using nominal modification. When considering a novel construction, speakers will then build upon their experiences with similar constructions, plausibly resulting in families of similar LVCs.

1.3.2 Construction Grammar

The idea that certain combinations of words carry meaning beyond what can be predicted by the component parts is an idea that has been recognized by Construction Grammar (see for example, Goldberg, 1996, 2006; Croft & Cruse, 2004). Construction Grammarians often point to idioms, such as *kick the bucket,* as evidence that speakers must not only store individual words and their meanings, but also certain phrases or "constructions" and their meanings. Traditional Generative Grammar largely treats such idioms as peripheral phenomena, which are part of the idiosyncrasies of a language (Chomsky, 1995). According to the theories of Construction Grammar, idioms are not peripheral or separate at all from other types of purely collocational language, rather there is a continuum from wholly fixed, non-compositional idioms like *kick the bucket*, in which no component part expresses the meaning of *die*, to semi-productive constructions like LVCs, in which the component parts do contribute lexical meaning and there is more flexibility as to what elements can form an LVC, to purely compositional or collocational language in which words combine freely and productively according to syntactic rules. This concept is displayed graphically in Figure 1.3.3.

Figure 1.3.2: Continuum of language compositionality.

14

1.3.3 Comparing Emergent Grammar and Construction Grammar

Both Emergent and Construction Grammar can be thought of as usage-based approaches in the sense that both schools assume that the structure of the language arises out of its usage. There are certainly a variety of somewhat distinct viewpoints that fall into each tradition, perhaps especially so for Construction Grammar, which covers work that varies quite widely in theoretical specifics. The work of Adele Goldberg within Construction Grammar is, for example, quite distinct from the work of Ivan Sag on Sign-Based Construction Grammar. This research does not attempt to espouse one school over another, and rather draws upon ideas emphasized in both, including the importance of frequency in grammar.

It has been argued that Construction Grammar makes certain claims on the storage and processing of lexical items: specifically that phrasal constructions, as pairings of form and meaning, are stored in the mental lexicon in the same way that individual lexical items are stored in the lexicon, suggesting that constructions are not decomposed or analyzed compositionally (Piñango, Mack & Jackendoff, 2006; Wittenberg & Piñango, 2011). Whether or not all or most Construction Grammarians would agree with this viewpoint is debatable (Kay & Sag (submitted), for example, develop a lexical theory of phrasal idioms within the framework of Sign-Based Construction Grammar). If compositionality is graded, then it follows that perhaps some constructions are stored as unanalyzed wholes, while others are understood in both their compositional and idiomatic, non-compositional interpretations. This research explores some of these issues of storage and processing of constructions, and does not assume that all constructions are stored and processed as if they were single lexical items.

Chapter 2
Defining & Delimiting LVCs (Research Question 1)

One of the clearest challenges to the study of LVCs is determining a definition and delimitation

of what is to be studied. This is not a simple issue because LVCs exist on a continuum from

purely compositional language on one extreme and entirely non-compositional, opaque idiomatic

expressions on the other extreme (see Figure 1.3.3). LVCs can also be thought of as lying on a

continuum of verbal elements - from fully predicating "heavy" verbs on one end of the spectrum

to non-predicating auxiliary verbs on the other (a graphical representation, similar to that

provided for the continuum of compositionality, is provided below in Figure 2). As a result,

LVCs can often overlap with any of these other elements on either spectrum.

Aspectual Verbs:
*She **began** to cry.*
&

| "Heavy" Usages: | Causative Verbs: | Semi-Light Verbs: | LVCs: | Auxiliaries: |
| *I **gave** him a present* | *We **grant** him control* | *I **gave** them a hint* | *She **took** a walk* | *We **did** go.* |

High Predicational Power ← Continuum of Verb Types → No Predicational Power
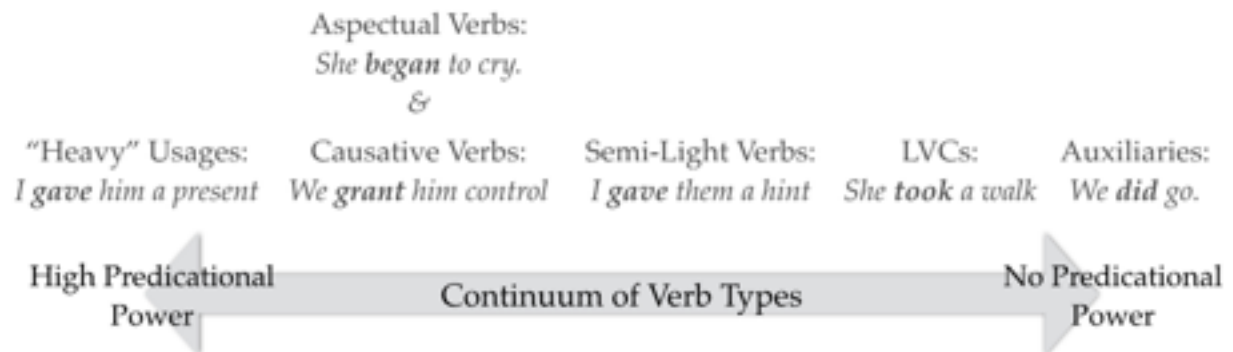
Figure 2: Continuum of verb types, based on the type's predicational power.

A discussion of linguistic approaches to distinguishing LVCs from auxiliaries, raising

constructions with two predicates, idiomatic expressions, and heavy verb constructions follows

in Section 2.1. In Section 2.2, an overview of computational approaches to detecting and

delimiting Multi Word Expressions and LVCs is provided. Section 2.3 shifts to original work:

the development of the PropBank annotation procedures for identifying and annotating LVCs is

described, including comparisons to other annotation schemas for LVCs and other support verb types seen in Figure 2. The utility of these annotations is demonstrated through the creation of the state-of-the-art automatic detection system that was developed using these annotations as training data; a description of this system is given in Section 2.4.

2.1 Linguistic Approaches to Defining and Delimiting LVCs

2.1.1 Distinguishing LVCs from Auxiliary Verbs

Within the minimalist program, light verbs are treated as instantiations of *v* (Adger, 2003: 134), and therefore given a unique syntactic representation. However, this treatment tends to conflate light verbs with auxiliary and modal verbs, as well as other elements that are not verbs, or in some cases, are not overtly expressed, which are also captured by *v*. This conflation is not uncommon. Because light verbs are thought to contribute a limited quantity and quality of semantics to LVCs, they have often been seen as a type of auxiliary verb. Auxiliary verbs, somewhat like light verbs, are subordinate to a main predicate (verb) but do contribute a certain set of semantic features, such as mood, aspect and voice (Crystal, 2003). Given this similarity, several theories group LVCs with auxiliary verbs (Hacker, 1958; Hook, 1974, 1991, 1993; cited in Butt, 2003). Similarly, both Abeillé, Godard and Sag (1998) and Rosen (1989) develop analyses that group light verbs with tense auxiliaries and causative constructions in Romance languages. These analyses are compatible with the diachronic views of auxiliaries, wherein full verbs gradually lose their meaning. Hopper and Traugott (1993) include light verbs, which they

term "vector verbs", as an optional stage in the grammaticalization cline, moving from full verb, to vector, auxiliary, clitic and finally affix.

However, as Butt (2003) points out, there are several important ways in which light verbs differ dramatically from auxiliaries. First, if light verbs were truly part of a language's set of auxiliaries, we would expect there to be some restrictions on the interaction of light verbs with auxiliaries, as we see with the established auxiliary verbs of the language. For example, while English main auxiliaries *do, be, have* can interact with modal auxiliaries *can/could, may/might, shall/should, will/would,* etc. (e.g. *She should have gone*), there are restrictions on the co-occurrence of main auxiliaries with other main auxiliaries and modals with other modals in most English dialects (**She may should have gone*). There are no such restrictions in the interactions of light verbs and auxiliaries; rather, light verbs can co-occur with the full paradigm of auxiliaries (Butt & Geuder, 2001): *She should have made an appearance*. Light verbs also do not behave in a fashion that is similar to auxiliaries as far as a special negative form (e.g. *isn't, wouldn't*) and subject inversion (e.g. *is he, will they, *made she an appearance*) (Crystal, 2003).

Additionally, although the level and quality of semantic content that light verbs contribute can vary across languages and even across light verbs within the language, the semantic contribution of light verbs always has the potential to go beyond that of tense, aspect and mood information. Crosslinguistically, we see light verbs contributing volitionality, forcefulness and benefaction (Butt & Geuder, 2001). In English, the semantic contribution is clear when comparing pairs of light verbs, as seen previously: *She took a bath* is clearly very different from *She gave a bath* in a manner that is quite distinct from a difference of auxiliaries. Using *give* instead of *take* adds an argument to the subcategorization frame of the bathing event.

Finally, light verbs remain identical in form to main verbs, rather than diminishing in form as auxiliaries do. English light verbs have existed since Old English, and yet have remained relatively stable in form (for a further discussion of light verbs in Old English, see 3.2). Thus, outside of what seem to be some incidental similarities to auxiliary verbs, there is evidence and criteria for distinguishing light verbs from auxiliaries.

2.1.2 Distinguishing LVCs from Two-Predicate Constructions

Having established that LVCs are distinct from auxiliary verbs, let us now consider what definitional criteria can distinguish light verbs from control and raising constructions wherein two predicates are present, as opposed to a single complex predicate. Although Butt (2003) encourages syntactic criteria for LVCs that are language-dependent, she points out a useful syntactic feature that seems to be valid across light verbs in all languages: monoclausity. Monoclausity is simply the property describing any construction containing only one clause. With monoclausity as a definitional criteria, it is possible to distinguish light verbs from control and raising constructions, which, in Butt's theory, involve clausal complements. Furthermore, this requirement underscores an important characteristic of LVCs, which is that the grammatical structure is essentially that of a simple predicate, but there are two or more heads contributing arguments as part of the primary predication (Butt, 2003). This interpretation is somewhat at odds with some research, which groups together control and raising verbs with light verbs (e.g. Rosen, 1989); nonetheless, it is an important and clear feature that distinguishes the syntactically and semantically distinct control and raising structures from LVCs in English.

Van Valin and LaPolla's (1997) work on complex sentences and clause juncture provides some evidence for the interpretation of LVCs as monoclausal structures (441-448). Van Valin and LaPolla's theoretical framework is Role and Reference Grammar, and involves syntactic notions of a clause consisting of a "core" and "periphery." A core consists of at least one "nucleus." The periphery is defined as the elements of a clause that are not arguments of a predicate (e.g. locative phrases, *in the library,* and temporal phrases, *at noon*). The core is defined as the predicate and its arguments, and the nucleus is defined as the predicate. A graphical representation of this theory of the clause is given in Figure 2.1.2.

CLAUSE

CORE

PERIPHERY

NUCLEUS

John **saw** Mary

in the library.
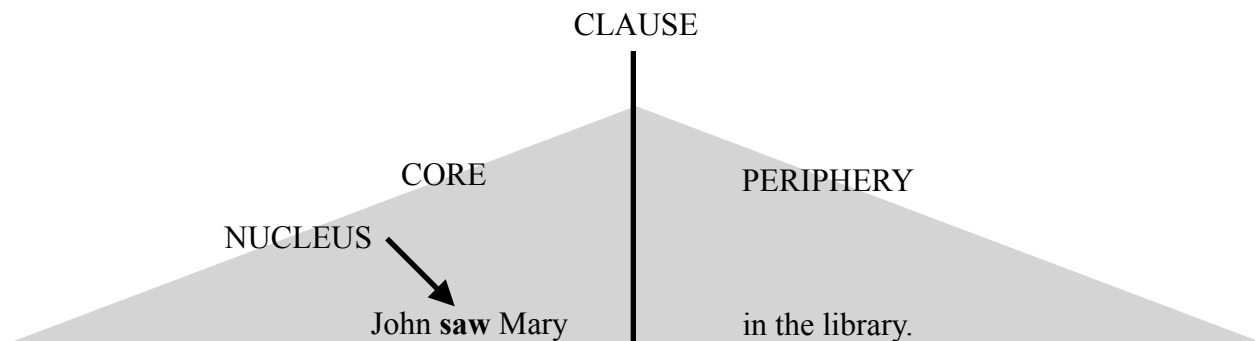
Figure 2.1.2: Van Valin & LaPolla's (1997) clause structure.

The authors discuss three distinct patterns of these components in complex sentences: (1) nuclear junctures involving a single core containing multiple nuclei (e.g. *John forced open the door*), (2) core junctures involving what is postulated to be a single clause made up of multiple cores (e.g. *Mary persuaded Sally to leave*), and (3) clausal junctures involving a single sentence made up of two clauses (e.g. *Mary called Fred yesterday and asked him to paint her room white*). Of interest for this research are the diagnostics the authors discuss to distinguish nuclear juncture from core juncture. This is of interest regarding LVCs because nuclear juncture involves two nuclei, or predicates, acting as a single complex predicate and sharing arguments, while core juncture

involves two nuclei that are part of two distinct cores with both shared and distinct arguments of each core (for example, the shared argument in *Mary persuaded Sally to leave* would be *Sally*). In terms of Butt's argumentation, nuclear juncture forms a complex predicate like that of LVCs, while control and raising structures would be instances of core juncture, with nested but distinct predicate argument structures.

The first key difference between nuclear and core juncture discussed by Van Valin and LaPolla (1997) is the level of cohesion between the two nuclei, demonstrated by what elements, if any, can intervene between the two predicates. Two nuclei may be adjacent in some nuclear junctures, whereas the two nuclei cannot be adjacent with core juncture, and generally require a complementizer, often in the form of infinitival *to* in English: *Mary persuaded Sally to leave,* **Mary persuaded Sally leave*. If this test is applied to LVCs, it is clear that the two nuclei are generally either adjacent or separated only by determiners and adjectives: *Mary **took advantage** of the situation.* It is also clear that the two predicates in LVCs do not require an intervening element as instances of core juncture do.

The authors also use reflexivization as evidence for the presence of one or two cores, since reflexivization is only possible for syntactic co-arguments within a single core. Thus, the authors point out that in the case of core juncture, involving two cores, reflexivization is impossible: **Fred asked Pam to help himself.* Again, if we attempt to apply this test to LVCs, reflexivization seems possible, indicating that LVCs are instances of nuclear juncture: *He is only **taking advantage** of **himself**.* However, this test is somewhat problematic when applied to constructions with nominal predicates, especially if one postulates empty subject arguments with such predicates, as is sometimes done with verbal counterparts. For example, in the case of *Fred*

wants to help himself, the reflexivization is thought to be possible because *himself* is plausibly coreferential with the null subject of *help*. One could argue that nominal predicates also have such coreferential null subjects, and that is why reflexivization is possible, but this still leaves open the possibility of two distinct cores. Thus, it is a bit unclear how the reflexivization test can be fruitfully applied to LVCs.

Finally, the authors point out that in cases of core juncture, temporal modifiers can modify a dependent core independently of the matrix core: *Sam asked Fred to **leave tomorrow***. In the case of nuclear juncture, since the two nuclei are acting as a complex predicate, temporal modifiers must apply to both nuclei. This seems to be the case with LVCs: *Mary made an offer yesterday* cannot be interpreted to mean that the *making* event and *offer* event are at separate times. However, if this test is applied to constructions with causal verbs and noun predicates, for example, *Sam caused the death of Fred today,* one may also problematically conclude that these are cases of nuclear juncture since it seems a bit strange to conclude that the *causing* happened *today,* but perhaps the *death* did not. Thus, Although Van Valin and LaPolla's diagnostics are applied to two verb predicates in their discussion, an attempt to apply the diagnostics to LVCs does provide evidence that LVCs are instances of nuclear juncture, rather than core juncture. This supports the interpretation of LVCs as constructions that involve a single, complex predicate, distinct from control and raising constructions. Nonetheless, it should be noted that applying such tests to constructions with nominal predicates is somewhat problematic.

2.1.3 LVCs and Idioms

Unlike auxiliaries, which have some clear diagnostic criteria, idioms themselves are somewhat

difficult to define because expressions exist along a continuum of idiomaticity. Generative

approaches to idioms have pointed to the non-compositionality of the expressions (e.g. Katz &

Postal, 1963: 275; Chomsky, 1980: 149), defining idioms as expressions whose meanings are not

predictable from the component parts. More recently, it has been demonstrated that while some

idioms are truly non-compositional, (e.g. *sawing logs, kick the bucket*), most are somewhat

compositional in that the meaning of the expression is metaphorically related to the meanings of

the component parts (e.g. *Spill the beans = reveal a secret*) (Nunberg, Sag & Wasow, 1994).

This evidences the notion of a continuum from purely compositional phrases to non-

compositional phrases. Within this continuum comes an accompanying range of syntactic

flexibility. What Nunberg, Sag & Wasow refer to as "idiomatically combining

expressions" (1994) are more syntactically flexible than the non-compositional idiomatic

expressions. For example, idiomatically combining expressions can undergo various syntactic

operations like passivization and topicalization, and certain components can be modified,

quantified or referred to anaphorically, pointing again to the existence of idiom chunks that

compose the phrase rather than a non-compositional expression (Nunberg, Sag & Wasow, 1994:

503). Non-compositional idiomatic expressions cannot undergo these operations: *The big

bucket was kicked by Ivan* has lost its relationship to dying in the transformation.

In light of these theories, it is clear that LVCs cannot fit the strict definition of a non-

compositional idiomatic expression. Most LVCs do undergo passivization, and LVCs modified

by adjectives and relative clauses actually seem to be more common than LVCs without

modification (Nickel, 1978; Brinton & Akimoto, 1999; Matsumoto, 1999). Kearns (2002) does

draw a distinction between "true light verbs" and "vague action verbs," adopted by some

researchers in computational linguistics (Stevenson, Fazly & North, 2004; Samardžić & Merlo,

2010). This distinction is also made on the basis of syntactic flexibility. True light verbs are

thought to be less compositional, and therefore less syntactically flexible. For example, some

LVCs are not generally compatible with the passive, referential determiners, or

pronominalization: *A groan was given by Sarah, ?Sarah gave that groan, *Which groan did

Sarah give? Kearns distinguishes these cases from "vague action verbs" which are more

compositional and therefore more syntactically flexible: A speech was given by the president,

The president gave that speech, Which speech did the president give? Semantically, however,

even true LVCs cannot be considered purely non-compositional since there is clearly a

relationship between the noun, groan, and the semantics of the expression.

It is not immediately clear how, or whether, one should draw a distinction between

idiomatically combining phrases and LVCs. There are certainly idioms listed in the work of

Nunberg, Sag & Wasow that fit the light verb + eventive/stative noun definition: make a dent in,

make a fuss over, make allowance for, make arrangements for/with, take a look at, take a try at,

take notice of, take pity on (1994: 533-534). Even in this short list, it seems that there is some

variability in the idiomaticity of the expression, and this variability seems to correspond loosely

to the expression's similarity to the lexical verb. For example, make a dent in does not

necessarily involve literal denting, instead it could expresses that little progress was made. On

the other hand, *take a look at* and *look* seem to have relatively less difference in meaning, and it seems reasonable to assume that *taking a look* involves literally *looking* at something.

Another shared characteristic of both idioms and LVCs is a resistance to substitution. Although there are certainly "families" of idioms that have similar meanings with substitutable terms (e.g *hit the sack/hay, pack a punch/wallop*) (Nunberg, Sag & Wasow, 1994:504), it is also undeniable that there are limitations and restrictions on substitution (*?hit the bed*). This is also true of LVCs, wherein some semantically similar substitutions of the noun are possible (*make a suggestion/recommendation*), but others are not (?*make advice*).

Given the clear overlap between idioms and LVCs, it seems plausible that there is a diachronic relationship between the two. Akimoto (1999) proposes that the noun complements in LVCs are originally concrete nouns characterized by all noun properties (articles, pluralization, modification), but with frequent use the noun takes on a stronger unity with the verb, and as this takes place, the verb too loses its more concrete meaning. This is evidenced when one examines *give* LVCs, which can lose compatibility with an in indirect object over time; for example, *give him a chase* of early Modern English was exclusively realized as *give chase* in late Modern English (Akimoto, 1999: 234). There are several problematic aspects of this argument, including the preponderance of articles, modifiers, and plurals in LVCs, which would indicate that nouns in LVCs do not necessarily become "decategorialized," as Hopper would put it (1991). Additionally, it is unclear how one could argue that some of the nouns in LVCs were ever concrete in any sense. Nonetheless, it seems certain that there is overlap between idiomatically combining phrases and LVCs, but given that the two are not at odds and both are

rather loosely defined, perhaps this is not surprising. It seems most precise to describe LVCs as a subtype of idiomatically combining phrases, which have special constraints of their own.

2.1.4 Syntactic Approaches to Distinguishing LVCs from Heavy Verbs

Distinguishing LVCs from syntactically identical heavy *verb + noun* phrases is perhaps one of the most important and challenging tasks for LVC identification. Butt & Geuder admit that there do not seem to be any clear syntactic criteria for distinguishing English light verbs from main verbs; for example, the commonly cited English light verb *give* is found in the same ditransitive syntactic frame as the heavy, transfer sense of *give* (2001: 339). Consider the following syntactic trees:
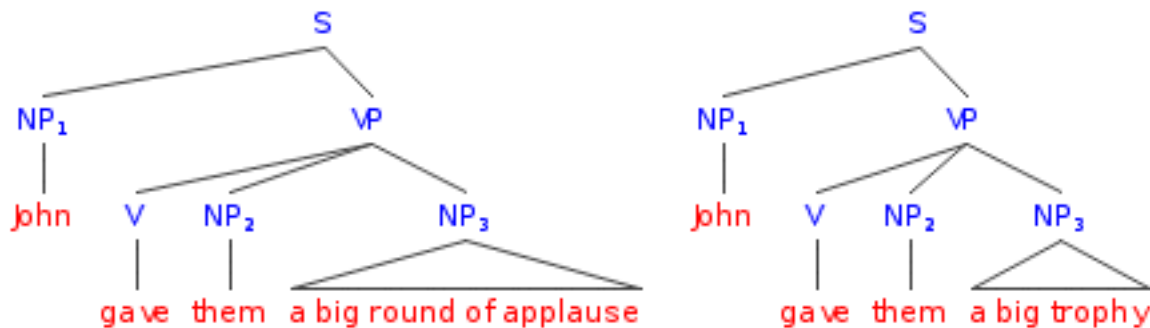
Figure 2.1.4-1: Syntactic trees for light (left) and heavy (right) usages of *give*.

In some cases, it seems that argument attachment could illuminate the distinction between light and heavy usages of other English light verbs. For example, in the case of *take*, a goal argument would likely be a sister to the verb in heavy usages, but embedded in the object Noun Phrase (NP) of light usages:
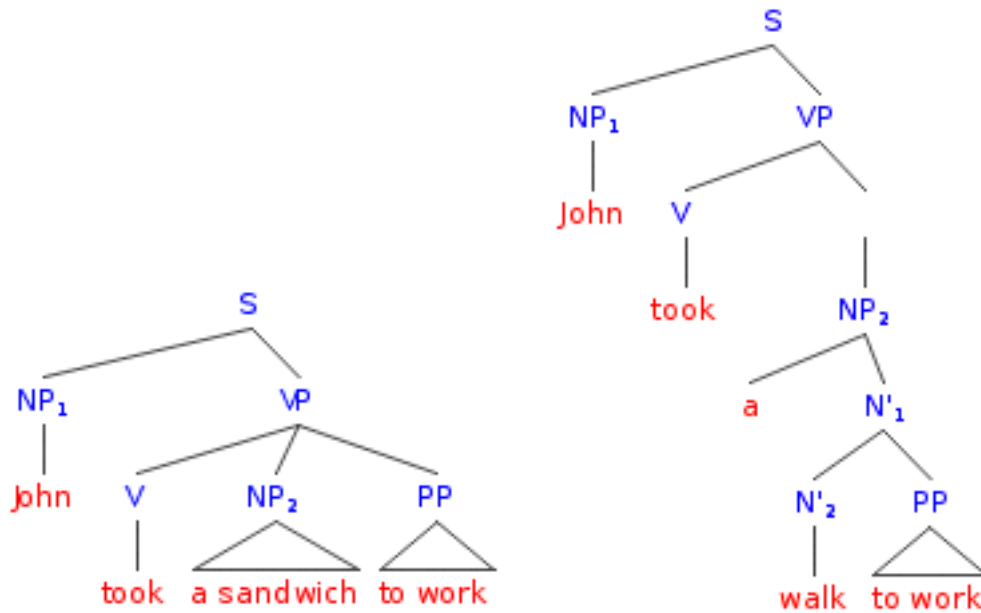
Figure 2.1.4-2: Syntactic trees with heavy (left) and light (right) usages of *take*,

Thus, although the syntax of English LVCs can sometimes be a clue to distinguishing between

light and heavy usages, it seems that other English LVCs are syntactically identical to heavy

usages (as seen in the *give* examples). Compounding this difficulty is the inherent circularity

involved in the fact that the creation of a syntactic tree such as that given for *John took a walk to*

work, with its distinct determiner and prepositional phrase attachment in comparison to *John*

took a sandwich to work, would rely on the creator of the tree having already decided that the

construction is a LVC. In cases where this distinction is not clear, the attachment of adjunct

arguments can be inconsistent, reflecting interpretations where the adjunct more directly

modifies the verb or the predicating complement. Consider a sentence such as *John gave them a*

round of applause for their efforts. Should the *for*-phrase be a modifier of *give,* or *applause*? As

Xue & Palmer (2009) indicate, whether or not a the predicating complement shares adjuncts with

a light verb is a difficult question; in general, the less semantic content the light verb has, the more likely it is that the adjunct is licensed by the predicating complement.

2.1.5 Semantic Approaches to Distinguishing LVCs from Heavy Verbs

Essentially, it does not seem plausible to define LVCs using purely syntactic criteria in English. Therefore, we will turn now to approaches that also involve semantic definitional criteria, beginning with a discussion of some theories on how semantic roles are assigned or projected by complex predicates like LVCs.

Grimshaw & Mester (1988) argue that light verbs are thematically incomplete, and therefore they assign no θ-roles (semantic roles, such as Agent, Patient, etc.) and impose no restrictions on the θ-roles of their syntactic arguments. Rather, the light verb bears inflection for the clause and assigns case, allowing the predicating complement to grammatically assign its own θ-roles in a verbal context. Although this notion of the predicating complement assigning θ-roles outside of its maximal projection defies some of the traditionally accepted notions of case assignment originally outlined by Chomsky (1981), Grimshaw & Mester (1988) propose a process of argument transfer, whereby arguments within the complement's constituent boundaries are θ-marked by the predicating complement, and arguments outside of that constituent, under S (the clausal node), are θ-marked by the light verb, which absorbs argument structure from the noun. In this analysis, the subject argument is always outside of the boundaries of the predicating constituent, under S. Cattell (1984) also examines what he calls 'complex predicates' within the theoretical framework of Chomsky's Government and Binding Theory, drawing similar conclusions about θ-assignment.

LVCs have also been analyzed as the unification of the lexical-conceptual event representations of the predicates involved (Butt, 1995; Broadwell, 2000; Rosen, 1989; Kearns, 2002), often assuming a framework similar to Jackendoff's (1990) conceptual semantics. Kearns (2002) uses a representation in which numbers are placed where arguments ordinarily sit for the heavy, main usage of a verb, but in the light usage, these numbers are place-holders that must be combined with the actual arguments (represented by letters) of another predicating element in order for the light verb to license actual arguments. Rosen embraces this representation, and gives the following example involving *give* (1989: 122-3). Rosen establishes that there are two representations for heavy and light *give* respectively:

heavy give (1x, 2y, 3z, E)
light give (1, 2, 3, E)

This representation indicates that heavy *give* licenses three arguments (Agent, Theme, Recipient) while light *give* only maintains the placeholders of the normal *give* subcategorization frame. For the following usages, Rosen (1989) argues that these placeholders combine with the arguments of *sweep* (which assigns an Agent sweeper and a Patient thing-swept) and *groan* (which assigns only an Agent groaner) respectively:

8. John gave the floor a sweep
9. John gave a groan

Light Representation for (8): give (1, 2, 3, E) + sweep (x, y, E) → (1x, 2y, 3, E)

Light Representation for (9): give (1, 2, 3, E) + groan (x, E) → (1x, 2, 3, E)

While such representations are intuitively and theoretically appealing in their explanatory power, the question remains as to whether argument structure patterning of this sort can be leveraged into statistical patterns that can be significant in the context of NLP. Nonetheless, these types of explanations of "argument sharing" are also being validated in cognitive work on the processing of LVCs, which will be discussed in Chapter 4 (Section 4.4.4).

Wierzbicka (1982) argues for the existence of strict semantic rules that govern the behavior of, and allow for the prediction of, constructions like *have a drink*. Wierzbicka does not call the focus of her research "light verb constructions;" instead, she focuses on what is arguably a narrow subset of LVCs, constructions of the form 'NP *have* + (auxiliary) + *a* + V-infinitive' (e.g. *have a drink*) (1982: 755). This definition excludes constructions like *have an argument/thought*, which involve a deverbal noun, and Wierzbicka also excludes some expressions in which the complement is a zero-derived nominal, and is therefore identical to an infinitival verb form. This is in part because, despite the presence of the indefinite article, Wierzbicka argues that the complements in the constructions of interest are infinitival verbs. She distinguishes form-identical deverbal nouns from verbs on the basis of stress and phonological features. Where these fail, she uses a test to determine how productive the pattern is, and whether it is productive with nouns or infinitival verbs. For example, she excludes expressions like *have a cough* based on the fact that there is a productive pattern requiring nouns that are semantically similar, denoting illness: *have a cold, have a fever*. From a practical perspective, making such a determination for other examples seems difficult. For example, should *have an*

30

answer be considered, although *have a response* exists and is semantically similar, although

clearly a noun? Does this constitute a "productive pattern" involving nouns? Besides this

difficulty, this analysis necessarily requires that one ignore what seem to be uncannily similar

phenomena. For example, Wierzbicka establishes a semantic formula for predicting this

construction in the context of joint speech activities (*have a chat, have a laugh*), but given her

narrow definition, would not consider the clearly related construction, *have a conversation*.

Precisely because LVCs do occur in semantically similar patterns, Wierzbicka's analysis seems

somewhat ill-suited to understanding LVCs in general.

Putting aside these difficulties for a moment, Wierzbicka goes on to argue that the *have a*

V constructions are not at all idiosyncratic, as they may seem, but rather are governed by

semantic rules. She postulates one general schema of the construction, specifying the semantics

of the construction in the semantic metalanguage of her own development (1972, 1980):

X had a V =
For some time, not a long time
X was doing something that could cause him to feel/know something
he was doing it not because he wanted anything to happen to anything other than himself
he could do it again (Wierzbicka, 1982: 758-9).

While Wierzbicka argues that this schema in itself has considerable predictive power, she admits

that it is not full predictive power, in that there are *have a V* constructions that would not fit this

general formula. Those constructions that do not fit require more nuanced semantic formulae for

different subtypes of the construction. Wierzbicka therefore develops 10 subtypes of the general

schema:

i.   Aimless objectless individual activity which could cause one to feel good (*have a walk, have a swim, run, job, lie-down*)

ii. Action aiming at perception which could cause one to know something and which would not cause one to feel bad if it didn't (*have a look (at), listen, smell, feel, taste (of)*))

iii. Tentative action which could cause one to come to know something and which would not cause one to feel bad if it didn't (*have a try, a look for, a think about*)

iv. Semi-Voluntary Action which could cause one to feel better (*have a cough, a yawn, a cry*)

v.   Consumption of small parts of objects which could cause one to feel pleasure (*have a bite, lick, suck, chew, nibble*)

vi. Consumption of non-discrete substances which could cause one to feel pleasure (*have a drink, a smoke, a sip, a sniff*)

vii. Activity superficially involving another entity, which could cause one to feel pleasure (*have a kick of the football, a throw of the boomerang, a read*)

viii. Self-directed action which could cause one to look better (*have a wash, shave*)

ix. Joint bodily activity which could cause the people involved to feel pleasure (*have a kiss, a cuddle, a dance*)

x.   Joint speech activity which could cause the people involved to feel pleasure (*have a chat, a gossip, a laugh*)

Wierzbicka points to the analysis of the *have a V* construction as evidence that only a semantically based grammar can be truly generative and predictive.

Used in a practical NLP application, if one could automatically detect the appropriate subtypes, then this analysis could potentially assist in defining the appropriate inferences of each subtype. However, automatically detecting the subtypes accurately seems unlikely. Indeed, as previously mentioned, it seems that it would be difficult for a human to reproduce Wierzbicka's research and predict the same membership of *have a V* constructions. Not only is the definition of "infinitival verb" in these cases subject to interpretation, but the semantic metalanguage itself seems imprecise and open to very different interpretations. If humans would not make the same judgments consistently, one can be certain that a computational system cannot learn to make the distinction. Furthermore, although Wierzbicka points out that the traditional linguistic question,

'Can one say X' is not a reliable tool for investigating the semantic aspect of a grammar, this seems to be precisely her own method of investigation, and the semantic formulae that she creates seem circular in that they are designed to accommodate the set of phenomena decided upon in advance.

Nonetheless, semantic criteria can be used to define LVCs in English and other languages where syntactic criteria fail. We will now turn to original work examining such criteria. In English, the second predicational element is generally a noun, and semantic features of this noun might be the key to distinguishing light and heavy usages. The nouns in LVCs must be abstract, and must denote events (*make an offer*) or states (*have knowledge of*). Unfortunately, determining what nouns are abstract and denote events or states has its own definitional problems. However, in some cases the distinction between LVCs and corresponding "heavy" transitive usages of potential light verbs can be drawn based on the differences in inferences that must be made in relation to the LVC, in comparison to those that would be made in relation to a normal transitive usage. These inferences can be teased out through the types of questions that can grammatically be asked in response to an utterance, such as the following questions, developed to probe the nature of each common light verb:

10. I made a cake for her birthday. → What did you make it out of?
11. I made an offer to buy the house. → *What did you make it out of?

12. I gave $20 of my own money. → Who did you give it to?
13. I gave a groan. → *Who did you give it to?

14. I took a peach from the tree. → Then where did you put it/what did you do with it?

15. I took a walk. → *Then where did you put it/what did you do with it?

16. I had a jug of water. → Where did you put it?

17. I had a sip of water. → *Where did you put it?

Each of these questions is meant to probe the nature of the object: is it a concrete, tangible object (or an object whose existence continues before and/or after the event), or is it expressing an abstract, transitory event? Usages where the possible light verb is followed by a concrete object are not light usages of the verb. Usages where the same verb is followed by a direct object expressing a transitory event are potentially light verb usages.

In some cases, there is necessarily overlap due to objects that can be construed either as concrete entities or as events:

18. I took a drink. → Where did you put it?

In these cases, additional context, including modifiers, can help to disambiguate heavy and light usages:

19. I took a long drink. → *Where did you leave it?[4]

20. I had a blue drink. → Where did you leave it?

_____

[4] The interpretation of a "long," as in "tall," drink is unlikely outside of contexts such as Las Vegas, where tall beverages abound.

Undoubtedly, other similar types of questions can be used, so long as the question probes the nature of the direct object. With a better understanding of the direct object, we can begin to distinguish between heavy and light usages of a potential light verb.

The key to implementing this distinction in a computational system firstly relies upon the existence of quality lexicons, like WordNet (Fellbaum, 1998), indicating what nouns potentially denote events and/or states. The presence of certain modifiers may also help to make a determination in borderline cases where nouns can represent either concrete objects or abstract events or states. For example, the temporal modifier in *I took a long drink* would give indication that the construction is a LVC, while the presence of certain adjectival and source modifiers in *I took a tall drink off the bar* would indicate that the construction is purely collocational. Thus, patterns of modifiers may provide crucial information for NLP systems to distinguish LVCs from other *verb + noun* combinations. For example, Fazly's (2007) research in detecting LVCs using features such as modifiers was quite successful, but was implemented on a relatively small scale.

To summarize, LVCs in some languages can be defined and delimited using clear syntactic criteria. Cross-linguistically, this seems to be the norm, as does complementary distribution between lexical verbs and LVCs for expressing certain semantic concepts. English and other Romance languages like French, however, defy the cross-linguistic tendencies in that a mixture of lexical verbs and LVCs exist together and seem to express very similar concepts, and the precise semantic difference between the two is very difficult to pinpoint. Additionally, in English, LVCs are syntactically indistinguishable from other *verb + noun* combinations. This leaves a primarily semantic definition of LVCs in English, and highlights the importance of modifiers in delimiting LVCs.

The following basic criteria define English LVCS:

1. monoclausal structure

2. verb of semantically general meaning

3. noun complement that is abstract, denoting an event or state

This basic definition is the starting point of annotation guidelines for PropBank LVC annotation, described in Section 2.3.

2.2 Computational Approaches to Defining & Delimiting MWEs and LVCs

2.2.1 Introduction to MWEs in NLP

Research in the field of Natural Language Processing (NLP) has, in the past, grouped LVCs into the larger category of Multi-Word Expressions (MWEs), all of which are quite challenging because MWEs often carry idiosyncratic meanings that cannot be fully predicted from the component parts. This includes a wide variety of constructions: compound nominals, such as *air conditioning,* fixed idiomatic chunks, such as *by and large*, and somewhat flexible verbal idioms like *toot/blow one's own horn*. Sag et al. (2002) is one notable work to recognize MWEs as a "pain in the neck" for NLP. Sag gives an overview of different varieties of MWEs, and discusses two basic approaches, or combinations thereof: the words-with-spaces approach or the compositional approach (2002). The words-with-spaces approach essentially treats all MWEs as if they were a single lexical item containing spaces; the compositional approach, at the other extreme, treats MWEs as equivalent to the single words that make them up (Sag et al., 2002).

These approaches can have varying levels of success, depending on what type of MWE they are being applied to, but overall each approach is problematic, and neither is sufficient for handling LVCs or other idiomatically combining phrases. The words-with-spaces approach is adequate for treating fixed expressions, such as *ad hoc* and *kingdom come*, because these expressions do not undergo any morphological or syntactic variation, nor any internal modification (e.g. **kingdom came*, **ad really hoc*) (Sag et al., 2002). Beyond fixed expressions however, the words-with-spaces approach quickly becomes less than ideal since all other types of MWEs undergo varying levels of syntactic and morphological variation and/or internal modification. Most LVCs are characterized by each of these: *he gives a speech* vs. *he gave amazing speeches*. Even in the case of Kearn's (2002) "true light verbs," which are less flexible syntactically, internal modification is possible: "She gave a little groan of dismay" (web example). In order to tackle such variety using the words-with-spaces approach, it would be necessary to have lexical entries that somehow captured each of these variations, not to mention an entry for every possible LVC, which is in itself a daunting task as there are literally tens of thousands of possible LVCs in English and likely more in other languages, such as Hindi or Jaminjung, the previously mentioned Australian language in which complex predicates are the majority (Butt & Geuder, 2001).

The compositional approach is perhaps even less desirable than the words-with-spaces approach, as it is problematic for all types of MWEs, including fixed expressions. The primary problem for this approach is the idiomaticity problem, closely related to which is the problem of overgeneration (Sag et al., 2002). The idiomaticity problem relates to the idiomatic nature of MWEs: if we are to treat, for example, the semi-fixed expression *trip the light fantastic* using the

compositional approach, one could account for the meaning of the MWE (to dance or move to music) only by assuming that there is a sense of each or some of these lexical items that somehow maps to the meaning of the expression (e.g. a sense of *trip* that means *dance*). This can result in an extreme proliferation of senses. Closely related to this problem is that of accounting for why this combination of words has meaning, but a combination of semantically similar words does not (**fall the soft fabulous*). LVCs are similarly problematic: the compositional approach does not provide an efficient methodology for classifying *he gave a speech* as an acceptable LVC, while rejecting *he gave a pontification,* nor the similarities and differences in meaning involved with LVCs such as *she took a bath* vs. *she gave a bath*. In a compositional approach, one may attribute this problem to the larger problem of selectional preferences generally: what makes certain objects compatible with certain verbs? This is a difficult question, but in the case of LVCs, the selectional preferences may be even more idiosyncratic and difficult to pinpoint than, for example, determining the selectional preferences of verbs such as *eat* and *drink.*

Nonetheless, there has been research attempting to pinpoint the features that make certain combinations of words compatible, while ruling out others. Wierzbicka (1982) is one example. In more recent work, Kay and Sag (submitted) develop a lexical theory of phrasal idioms, wherein they use Sign-Based Construction Grammar (SBCG) to provide the appropriate syntactic and semantic analysis of idioms, one in which a certain amount of syntactic freedom is allowed for certain idioms, but limited for others. The SBCG methodology requires that one postulate an idiomatic meaning for each idiomatic word, or combination of words (e.g. *beans = secrets* in *spill the beans*). Postulating such meanings may be more difficult on a large scale

(which may be required for NLP) than the authors indicate. Furthermore, it's not clear how this treatment would apply to LVCs. One option would be to assume that light verbs have figurative meanings, which would be both numerous and difficult to pinpoint. Another option would be to assume that light verbs are wholly semantically empty, and the semantics stem from the noun alone, which is essentially literal in its meaning. Thus, it's possible that LVCs would not be classed as idiomatic at all; if so, how would the differences in meaning across *give a bath* and *take a bath* be accounted for? Thus, we can begin to see some of the difficulties of implementing a compositional or lexical approach to MWEs on a large scale for the purposes of MWEs.

The problems associated with both approaches, words-with-spaces and compositional, strongly affect how LVCs and other MWEs should be represented in a lexicon. Large-scale lexical resources such as FrameNet (Fillmore, Johnson & Petruck, 2002) and WordNet (Fellbaum, 1998) continue to expand with entries that are MWEs (although FrameNet currently includes LVCs specifically as part of the annotation for the noun entry), representing the words-with-spaces approach. VerbNet is an online lexical resource that has recognized the problematic nature of representing verb behavior as if it were static when novel usages of verbs are constantly arising. One source of the extensibility of verbs is semantic and syntactic coercion, which allows a verb to be used in "atypical" contexts (Goldberg, 1995). For example, the typically intransitive verb *wiggle* can be used in a construction with an object, as in *She wiggled her foot out of the boot*, invoking the meaning of caused-motion (the Caused-Motion construction; Goldberg, 1995). As in the case of LVCs, constructional coercion can be considered "semi-productive" (Nickel, 1978), in the sense that the construction is compatible with a large variety of words, but these words must be semantically compatible in some respects for its use to be

extended to that context. Unfortunately, again, it is very difficult to pinpoint what aspects of

semantics make a word compatible or incompatible with a given construction. Thus, current

work on VerbNet includes efforts to use Hierarchical Bayesian Modeling (HBM) to capture

patterns of verb behavior, and therefore the statistical likelihood that a given verb will participate

in a given construction, including LVCs (Hwang, Nielsen & Palmer, 2010; Bonial et al., 2011).

Although it remains somewhat unclear how LVCs should best be represented in a lexicon, we

turn now to current approaches to detecting MWEs.

2.2.2 Detection of Multi-Word Expressions

Because MWEs do tend to have idiosyncratic meanings, it is important for Natural Language

Understanding systems to firstly recognize MWEs and avoid interpreting them compositionally.

Early efforts for identifying MWEs primarily used co-occurrence of component words: words

within MWEs tend to co-occur more frequently than they would if they were only used

compositionally (Manning & Schutze, 1999). Point-wise Mutual Information (PMI), was

established as a common measure of the strength of the association between two words (Church

& Hanks, 1990; Church et al., 1991), and therefore a good measure of the likelihood that a

collocation was, in fact, an MWE. PMI uses the probabilities of two words occurring

individually (for example, a given verb *v* and noun *n*), as well as their joint probability of co-

occurrence within a given window of words:

$$\mathrm{PMI}\,(v,\,n) \;\doteq\; \log \frac{P(v,\,n)}{P(v)\,P(n)}$$

This approach, however, does not necessarily distinguish compositional from non-compositional

expressions. To situate the problem within LVCs for a moment, the *verb + noun* combinations of

"buy shares" and "sell shares" have relatively high co-occurrence, but they do not constitute LVCs as they retain their compositional semantics (Grefensette & Teufel, 1995; Tan, Kan & Cui, 2006).

Lin (1999) uses PMI in a promising fashion to distinguish compositional and non-compositional usages. Lin compares the PMI of parts of an MWE with the PMI of a similar expression obtained by substituting one of the constituent words with each of a set of related words, drawn from an automatically created thesaurus developed by the author. The intuition behind this research is that the target expression is non-compositional if its PMI value is significantly different from that of any of its variants. Although intuitively appealing, Lin's implementation achieves rather low precision (the percentage of retrieved instances that are relevant) and recall (the percentage of relevant instances that are retrieved) of 15.7% and 13.7% respectively on a task classifying expressions as compositional or non-compositional.

However, we've established that compositionality is a continuum, so a binary distinction between compositional or non-compositional may be problematic. Venkatapathy and Joshi (2005) explored measuring the relative compositionality of collocations, thereby situating the expressions on a continuum of compositionality. The authors focused on *verb + noun* collocations, citing these as an especially important and challenging case to distinguish because some of these collocations may be used compositionally or non-compositionally interchangeably, depending upon the context. Potential LVCs are, of course, a good example of this property: *take a drink (off the bar)* ('grasp/move a beverage') vs. *take a drink (of the soda)* ('drink a beverage'). To measure the relative compositionality of such phrases, the authors collect a variety of statistical measures relating to a given collocation, including raw frequency and PMI,

and combine these measures with context-based features that use Latent Semantic Anlaysis (LSA) (Landauer, Foltz & Laham, 1998). LSA provides a representation by which one can compare the similarity of two contexts in which two words fall, and therefore a measure of the similarity of two words, assuming that they are similar in meaning if they frequently occur in similar contexts. Venkatapathy and Joshi use these measures as features combined into a vector that represents a given *verb, noun* collocation, and the resulting ranks of relative compositionality were compared to human ratings of compositionality, provided by two fluent English speakers. The agreement rates between annotators is somewhat low, between 0.61 and 0.71, measured using Kendall's Tau and Pearson's Rank-Order Correlations, respectively. The result of this effort is a 0.448 correlation, measured using Pearson's Rank-Order Correlation Coefficient.

McCarthy, Keller & Carroll (2003) focus on ranking the compositionality of verb particle constructions, such as *eat up*, which require a distinct interpretation from compositional usages of verbs followed by adverbial particles, such as *walk up*… The authors propose that such compositionality can be measured by comparing the neighbors of the lexical verb and the phrasal verb, as it is expected that the neighbors will be more similar for the lexical verb and compositional usages with adverbial particles (e.g. *walk, walk up*), while the neighbors will be less similar for the lexical verb and true verb particle constructions (e.g. *throw, throw up,* as in 'vomit'). For evaluation of the system, the authors obtain human rankings of the compositionality of a sample of 111 candidate phrasal verbs. The overlap of neighbors is calculated using a variety of different measures, and the resulting ranks from each measure are compared to the human ranking using the Spearman Rank-Order Correlation Coefficient. The

most effective measure involves counting the number of neighbors of the phrasal verb that have the same particle as neighbors of the corresponding lexical verb. This obtains significantly correlated rankings of the expressions to the human rankings.

Similarly, the research of Katz and Giesbrecht (2006) propose that compositionality can be measured by comparing the distribution vectors (based on LSA) associated with an MWE as a whole and those associated with its constituent parts. Greater similarity would indicate greater compositionality, while greater dissimilarity indicates non-compositionality. The authors test this proposal by attempting to determine when German expressions with identical surface forms are used compositionally, and when the expression is used figuratively or non-compositionally. The authors firstly show that the meaning vectors are quite different for a non-compositional MWE as a whole compared to its component parts. The authors secondly establish a threshold for how similar an MWE is to its component parts, under which the MWE should be considered non-compositional. The limitations to this approach are firstly that it relies upon there being a sufficient number of non-compositional usages in a corpus such that the meaning vector for the MWE reflects this usage, and it also assumes that the meaning of the compositional and non-compositional usages are highly distinct. If the meanings are somewhat related, local context measured by LSA will not be enough to distinguish the compositional from the non-compositional. Distinguishing surface-identical LVCs from heavy counterparts would likely suffer from this latter limitation, as an LVC such as *take a drink* surely shares meaning similarity with a heavy usage like *take a drink (off the bar)*.

While the above approaches take into account the lexical fixedness of many non-compositional phrases, they do not necessarily account for the relationship between syntactic

fixedness and non-compositionality. Fazly and Stevenson (2006) make use of this information, attempting to distinguish idiomatic, non-compositional *verb + noun* combinations from non-idiomatic, compositional combinations using measures based on both lexical and syntactic flexibility. The authors cite linguistic literature indicating that greater substitutability and syntactic flexibility of an expression indicates greater semantic analyzability and less idiomaticity (Glucksberg, 1993; Fellbaum, 1993; Nunberg, Sag &Wasow, 1994; cited in Fazly & Stevenson, 2006). Drawing upon this idea, they develop measures of both lexical and syntactic fixedness, and use these measures to distinguish compositional and non-compositional usages. Unlike Venkatapathy and Joshi (2005), in this research, the authors assume these usages can be discretely categorized. Fazly and Stevenson measure lexical fixedness in a manner inspired by Lin (1999), bringing together the PMI values of expressions varying by a single word (in this case, the noun of the expression) into a single measure that reflects the degree of lexical fixedness of the target *verb + noun* pair. Syntactic fixedness is measured by determining how often a target expression participates in the passive, which determiner types, and what variety of determiner types are used, and how often a noun is pluralized. Using these features to measure syntactic fixedness is motivated by the assumption that less compositional idioms are less likely to participate in passivization, certain determiner types will be canonical and others dispreferred, and pluralization is also either part of the canonical form (e.g. *spill the beans*) or not allowed (e.g. **spill the bean*). They develop a balanced data set of half idioms and half compositional usages, where idioms are identified in advance through the use of dictionaries with multi-word entries. Using PMI as an informed baseline of performance on this classification task, the authors find that their measure of lexical fixedness performs as well as this informed baseline,

and their measure of syntactic fixedness contributes to a 20% error reduction over the baseline.

When the measures are combined into a hybrid measure, there is another small improvement.

Fazly and Stevenson (2007) go on to develop a more comprehensive system that does

account for the continuum of compositionality. This research classifies *verb + noun*

combinations into subtypes of expressions with varying levels of compositionality: literal phrase,

abstract combination, LVC, or idiomatic combination. The authors use statistical measures to

quantify properties of these expressions that tend to gauge their level of compositionality, or

semantic idiosyncrasy. These include: measures of "institutionalization" (based upon the PMI of

the combination), measures of lexico-syntactic fixedness (based upon the measures of lexical and

syntactic fixedness discussed in Fazly and Stevenson (2006)), and measures of non-

compositionality (based on comparing the contexts of the expression with the contexts of its

constituents, like Katz and Giesbrecht (2006)). The classification of the automatic system is

compared to manual classification performed by a group of four annotators. Fazly and

Stevenson (2007) report F-scores (the harmonic mean of precision and recall) for the

performance of this classifier on each type of expression: the F-score is highest for classification

of LVCs (68%), somewhat lower for idiomatic expressions (56%), and lowest for abstract

combinations (46%). The authors point out that the measures of lexico-syntactic fixedness are

the most relevant for LVC identification.

Much of the research described here relies on the existence of manually annotated

training data. Such data can be both time-consuming and expensive to develop. Thus, Fazly,

Cook and Stevenson (2009) develop an unsupervised system for type and token identification of

idiomatic expressions involving "basic" verbs followed by noun phrases, such as *shoot the*

breeze. This system does not require manual annotation, relying instead upon automatically acquired statistical knowledge about idiom types, including measures of lexical fixedness and syntactic fixedness like those described in the previous paragraphs. The system builds upon the successful development of automatic NLP tools, such as syntactic parsers. Token identification of such expressions in context proves to be a difficult task. Focusing on detecting tokens of 51 expression types, the authors' system outperforms an informed baseline for expressions that have frequent literal usages, but does not outperform the baseline for expressions that are predominantly idiomatic. Overall, the accuracy of the system is comparable to other systems using a supervised approach.

In summary, we see a trajectory of MWE identification that begins with simple measures of the strength, or frequency, of co-occurrence of two words. More sophisticated measures firstly begin to account for the (dis)similarity of local contexts of the target expression, and then begins to also account for the relatively high syntactic fixedness of most non-compositional expressions. Future directions will likely include additional unsupervised approaches.

2.2.3 Detection of Light Verb Constructions

The research described in the previous section deals with detecting MWEs and non-compositional expressions generally, which does often entail the detection of LVCs. In this section, we will explore research in NLP that has been focused specifically on detecting LVCs. Some research involves LVC detection as an intermediate goal for the purpose of studying LVC productivity and the acceptability of certain *light verb + noun* combinations. This research will

be discussed in more detail in Section 4.2, which gives an overview of computational approaches to understanding LVC productivity. In this section, research on classifying LVC types and tokens in context will be the focus.

Some research in this area has relied upon features of the noun complement to identify potential LVCS. Grefensette and Teufel (1995) try to distinguish LVCs from other *verb + noun* combinations by focusing on the following features of the noun: the authors firstly expect that the noun is a deverbal noun, and secondly require that the noun shares similar argument and adjunct structures with its verbal counterparts. Verbs that frequently appear with nouns of this type are deemed light verbs. Dras and Johnson (1996) expand upon this approach, but also consider the verb's corpus frequency independently, which improves LVC identification because it models the fact that some verbs tend to be used more often as light verbs. Both of these approaches suffer from the inclusion of *verb + noun* combinations that are not light but simply occur frequently (e.g. "buy a share").

Tan, Kan and Cui (2006) build upon the work of Grefensette and Teufel (1995) as well as Dras and Johnson (1996), taking a supervised machine learning approach to the problem of classifying a combination as LVC or not LVC: each *verb + noun* combination constitutes an individual classification instance, where each combination possesses a set of features, and each is assigned a label for binary classification. The features used by the authors include some that are based on the work of Grefensette and Teufel as well as Dras and Johnson, but also include new features such PMI, comparing the frequencies of deverbal nouns to those of their verbal counterparts (assuming these should be similar in the case of LVC if LVCs and verbal counterparts interchange freely), and a "light verb class" feature, which amounts to restricting

possible LVCs to combinations involving the verbs *do, get, give, have, make, put,* and *take.* The authors test their system on the Wall Street Journal section of the Penn Treebank, from which they create a data set of 2,840 *verb + noun* combinations. Tan, Kan and Cui identify LVCs through manual annotation procedures wherein the authors themselves classify combinations according to two "levels of lightness:" Strict or Lenient. The Strict assignment is given where the median response from the three annotators is "Yes, it's an LVC." The Lenient assignment is given where the median response is either "Yes," or "Not sure." This type of classification seems to compare levels of certainty, as opposed to "levels of lightness," as the authors claim. The annotation examples provided by the authors are all arguably light or semi-light, adding to the somewhat confusing picture of the annotation schema — "take a chance" (arguably an LVC with a meaning akin to the verbs *risk* or *chance*) is given as the example of a "No" response. The authors achieve an F-score of 57.6% for the identification of Strict LVCs, and 68.9% for Lenient LVCs. The authors find that the "light verb class" feature gives the biggest boost in performance, which is both unsurprising and somewhat problematic in that this feature constrains and simplifies the problem greatly by limiting the scope of LVCs that will be detected to a limited set of verbs, when a larger variety of verbs can actually serve as light verbs in English (e.g. *keep track*)

Tu and Roth (2011) examine the state of the art surrounding MWE and LVC detection and note that there are two basic approaches: contextual and statistical. Statistical features are numerical features computed globally over large corpora, while contextual features are local features generated directly from input sentences. The authors further note that each of these approaches involve some overlapping tools or techniques, so they look to examine which

approach is the most useful by implementing LVC classification systems that use either

contextual features or statistical features, and then a combined approach. The statistical features

that they use are the PMI of a particular *verb + noun* collocation, the ratio of a deverbal noun's

usages to the number of its counterpart verb's usages, and the phrase size of the construction

(expecting LVCs, including modifiers and articles between the verb and noun, to be from 2-6

tokens, and not longer). The contextual features that they use are the identity of the head noun,

the bigram of *verb + noun*, Levin classes (which distinguish types of verbs with deverbal noun

counterparts that tend to co-occur with a given light verb, e.g. nouns related to sound emission

verbs: *make a clap, pop, whistle*), and a variety of other features such as the part of speech of the

preceding and following words and the determiner type.

Tu and Roth develop a balanced (approximately half LVCs, approximately half heavy

usages of the same verbs) data set generated from the British National Corpus (BNC).[5] They

first constrain their LVC data to *verb + noun* combinations that involve the most common light

verbs: *do, get, give, have, make* and *take.* They further filter the data by including only

combinations with head nouns that are either zero-derived nominals (e.g. *offer*) or derivationally

related to a verb according to WordNet. The authors then exclude combinations where the noun

is used less frequently than its related verb counterpart in the BNC. The authors estimate that

this rules out about 55% of the potentially negative examples, but it also surely rules out a

variety of LVCs that do not fit into the constrained category of combinations. The remaining

combinations are manually annotated by 2 annotators who volunteer for and complete the task

via a web interface. The "detailed" instructions for identifying LVCs in annotation are actually

---

[5] http://www.natcorp.ox.ac.uk/XMLedition/

quite sparse, and essentially just indicate that if an expression can be substituted by the verb related to the deverbal noun in the construction, then it should be classified as an LVC. Annotators then read sentences and indicate whether the usage is or is not an LVC, or if they are not sure. Tu and Roth then accepted all annotations where the 2 annotators agreed on LVC-hood, resulting in a data set of 1,039 positive examples and 1,123 negative examples.

The authors find that a combined system, using both contextual and statistical features, achieves an F-score of 86.3%, well over a (chance) baseline of 52.2%. However, the authors find that the two types of features (statistical and contextual) actually represent very similar knowledge, therefore combining them doesn't provide very much additional information. Only when the data set is reduced to cases of *verb + noun* combinations where the surface form is identical for an LVC and a non-LVC usage do the contextual features become more useful. This is because the contextual features take into account features of surrounding words, while statistical features treat identical collocations identically. Tu and Roth cite an example of syntactically identical collocations:

21. He **had a look** of childish bewilderment on his face.

22. I've arranged for you to **have a look** at his file in our library.

Here, the first sentence is not an LVC, and is closer to the meaning "possess an expression," while the second sentence is an LVC, meaning "look at." In these cases, it is clear why contextual features such as the following preposition, "at" or "of," can provide important disambiguating information.

2.3 The Development of PropBank LVC Annotation

The previous sections have illustrated the importance of quality manual annotations of LVCs to serve as the training data for automatic detection systems, and such annotations are also important for advancing theoretical studies of LVCs. In most cases, existing manual annotations are constrained in ways that do not recognize the full spectrum of LVCs by focusing only on LVCs that involve certain verbs, certain nouns that are derivationally related to verbs, or certain phrase types (e.g. *verb + noun phrase*, which rules out LVCs that involve prepositional phrases, such as *take into account*). Turning now to original research, the following sections describe our development of an annotation schema for recognizing and annotating the predicate argument structures of LVCs for the PropBank corpus (Palmer et al., 2005).

2.3.1 PropBank Background

The primary goal in developing the Proposition Bank, or PropBank, was the development of an annotated corpus to be used as training data for supervised machine learning systems. The first PropBank release, PropBank I, consists of one million words of the Wall Street Journal portion of the Penn Treebank II (Marcus, Santorini & Marcinkiewicz, 1993). PropBank provides annotations of predicate-argument structures for verbs and other relations, using semantic role labels for each argument. Although the semantic role labels are purposely chosen to be quite generic and theory neutral, Arg0, Arg1, etc., they are still intended to consistently annotate the same semantic role across syntactic variations. Thus, the Arg1 or Patient (shown in bold-face) in "John broke **the window**" is the same window that is annotated as the Arg1 in "**The window** broke," although it is the syntactic subject in one sentence and the syntactic object in the other.

The resulting PropBank annotation supplies consistent, simple, general purpose labeling of

semantic roles for a large quantity of coherent text to support the training of automatic semantic

role labelers, in the same way the Penn Treebank has supported the training of statistical

syntactic parsers.

PropBank provides a lexical entry for each broad meaning of every annotated verb,

including the possible arguments of the predicate and their labels (its 'roleset') and all possible

syntactic realizations.[6] For example, the verb *leave* includes the following two rolesets, which

correspond to syntactically and semantically distinct senses of the verb:

Roleset ID: leave.01 *move away from*
Roles:
Arg0: entity leaving
Arg1: place, person, or thing left
Arg2: attribute of arg1
Example: *John left Mary alone.*

Roleset ID: leave.02 *give*
Roles:
Arg0: giver/leaver
Arg1: thing given
Arg2: benefactive, given-to
Example: *Mary left her daughter the diamond pendant.*

This lexical resource is used as a set of verb-specific guidelines for annotation. All of the senses,

or rolesets for a given relation are compiled into a Frame File. In addition to numbered roles,

PropBank defines several more general (ArgM, 'Argument Modifier') roles that can apply to any

verb, which are similar to adjuncts. These include LOCation, EXTent, ADVerbial, CAUse,

---

[6] The PropBank lexicon of rolesets can be found here: http://verbs.colorado.edu/propbank/framesets-english/

TeMPoral, MaNneR, and DIRection, among others. These are marked, for example, as 'ArgMLOC.'

In the past, PropBank annotation had been restricted to verb relations, but more recent work has extended coverage to noun relations and complex relations like LVCs. Previously, the importance of the noun in such constructions was handled separately by NomBank (Meyers et al., 2004), as only verb predicates were annotated in PropBank. Light verb usages were handled either by simply being lumped in with one of the most dominant, semantically general senses of the verb (e.g. make.01, the creation sense of *make*), or through the designation of a roleset listing specific constructions (e.g. *make a bid*). This practice precluded delving into the deeper semantics of these constructions, represented by the noun as well as the verb.

Because annotation efforts had originally focused on only verb relations, this also left a gap in the coverage of events that can be expressed as noun relations, adjective relations, or as MWEs like LVCs. Within a language and across languages, the same event can be expressed with different syntactic parts of speech, for example:

23. He fears bears.

24. His fear of bears…

25. He is afraid of bears.

Or, for example:

26. He offered to buy the house.

27. His offer to buy the house…

28. He made an offer to buy the house.

Thus, it has been necessary to expand PropBank annotations to provide coverage for noun, adjective and complex predicates (Bonial et al., 2014a). An effort was made to restrict the annotation of nouns to eventive nouns, and pre-existing NomBank Frame Files were used as extensively as possible.

Expanding semantic annotation to other predicate types also requires adaptations in how the annotations are done. To best leverage the syntactic annotations upon which PropBank is annotated, PropBank verbal annotation uses the Penn Treebank syntactic annotations (Marcus, Santorini & Marcinkiewicz, 1993) to define the domain of locality for a predicating element, and only those constituents within the domain are annotated as arguments. While defining the relevant scope of annotation is a vital part of the PropBank annotation procedures, this is extremely dependent upon the syntactic characteristics of the predicate, and therefore procedures had to be adapted for LVCs. The annotation of complex predicates, such as the LVC *make an offer*, calls for annotation of the syntactic arguments of both the light verb *make* and the noun predicate *offer* (for a full description of the development of PropBank LVC annotation guidelines, see Hwang et al. (2010a)).

Another difficulty with such expansion is the efficient creation of Frame Files, which house the possible senses or rolesets for a given relation. Because Frame Files have been tied to a particular lexical item and its part of speech, moving to new predicate types necessitates the creation of hundreds of new Frame Files. This process can be extremely time-consuming, but its importance cannot be underestimated. The following sections discuss the challenges of Frame File creation and annotation for LVCs.

2.3.2 PropBank LVC Annotation Schema: Logistics

The first difficulty concerning the development of an annotation schema for LVCs involves a basic problem already mentioned in the realm of NLP: how do we account for LVCs, and the productivity of LVCs in a lexicon? LVCs are productive enough that a words-with-spaces approach, which would require the creation of a frame file for each unique LVC type (i.e. a frame file for *make_offer, take_bath, get_slap*, etc.), is not tractable. Nonetheless, annotators need a roleset in order to provide the appropriate numbered arguments. Since the bulk of the semantics in an LVC stem from the noun, we decided that LVC annotation could make use of the noun frame files, the creation of which was bolstered by the existence of NomBank Frame Files. The NomBank rolesets were originally developed to be parallel to verb rolesets where a noun and verb predicate were etymologically related. Thus, for example, the rolesets for *destroy* and *destruction* are identical:

Arg0: *destroyer*
Arg1: *thing destroyed*
Arg2: *instrument of destruction*

Currently, these rolesets would also be mapped via a Source field in the roleset, indicating that the two rolesets are related to one another. Building upon the NomBank resources, PropBank continued to add noun rolesets semi-automatically by simply copying the roleset of an etymologically related verb if it existed. Of course, in some cases, this was problematic because occasionally noun relations are characterized by arguments that aren't grammatical with the related verb relation, or vice-versa. For example, the noun *profit* is often realized with an argument indicating the amount of profit:

55

29. A profit of $20…

The verb *profit* is rarely realized with this argument, and is instead often realized with the source

of profit:

30. ?He profited $20.
31. He profited from the sale of stocks.

In these cases, the rolesets of related verb and noun relations were allowed to differ with respect

to arguments unique to the relation's part of speech.

While this solved the problem of providing rolesets for guiding LVC annotation, the

issues of the annotation task pipeline, annotation span and the final representation remained. In

PropBank practices, each annotation instance contains a pointer to a single relation for

annotation, and the corresponding frame file is displayed for the annotator to select a particular

roleset. Although certain verbs tend to serve as light verbs, there is a wide variety of verbs that

can be used in LVCs in lower frequencies. To capture the full range of light verbs, the roleset

option 'LV' was added to all rolesets, allowing annotators to indicate that a usage is light for any

verb relation. If annotators decided that a usage was light, they were to select the LV roleset and

mark the predicating noun relation with a special tag: PRR (PRedicating Relation). This is

termed the "verb pass" of LVC annotation. A screen shot of the PropBank annotation tool,

Jubilee (Choi, Bonial & Palmer, 2009) is given in Figure 2.3.2-1. Notice that in this case, the

verb *do* is marked as the relation ('rel'), but the annotator has selected the LV roleset (do.LV at

the top right) and marked the noun of the LVC, *search* as the true predicating relation ('ARGM-

PRR' — note that the ARGM notation is used for convenience, but these are not treated as

modifier arguments).
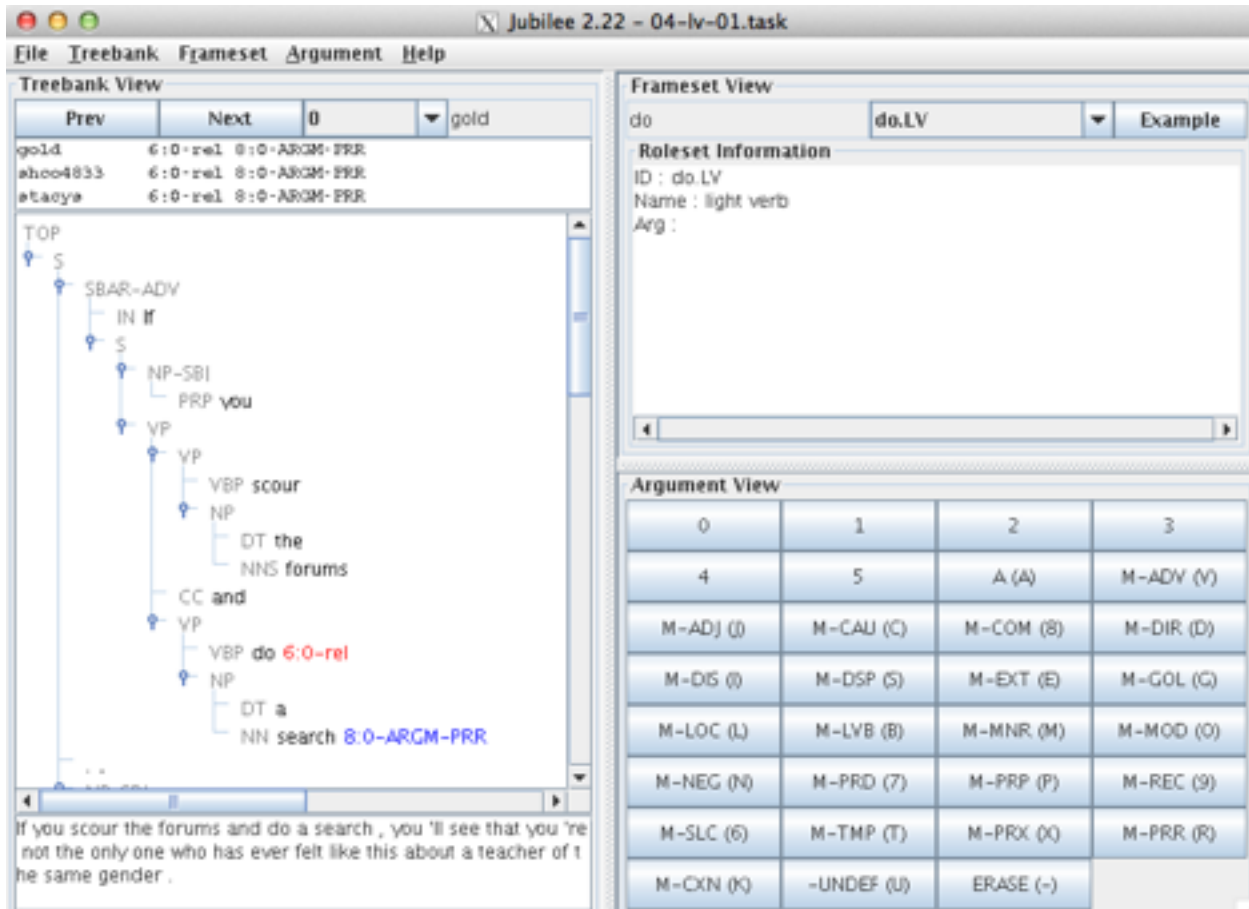
Figure 2.3.2-1: PropBank's annotation tool, Jubilee, displaying the verb pass of LVC annotation.

Thus, the verb pass of annotation allows for annotators to recognize the light usage of any

verb, and mark the true noun predicate. To capture the full semantics of the LVC and mark its

predicate argument structure, all instances marked as LV undergo a second pass of annotation,

wherein the lexical item marked as PRR in the first pass is then marked as the relation. In this

annotation pass, the syntactic arguments of both the light verb and the predicating noun are

annotated according to the noun's roleset. A screen shot of the corresponding second pass of

annotation is given in Figure 2.3.2-2. Notice that in this pass, the noun is marked as the relation,

while the verb is marked as a light verb with the notation LVB, and the annotator has assigned

the Arg0 role to the verb's subject, according to the roleset for the noun, search.01, displayed in

the top right.

Figure 2.3.2-2: PropBank's annotation tool, Jubilee, displaying the second pass of full LVC annotation.

To account for the fact that LVCs are truly complex predicates, and that *giving a bath* is distinct

from *taking a bath*, the final representation collapses the noun relation and LVB into a single

complex relation. These annotations are stored as text strings showing the arguments assigned to

particular nodes in the tree, along with a pointer to the appropriate tree. For example, in this case

the stored annotation would be: `gold 1:1-ARG0 6:0,8:0-rel`, thereby marking both *do* and

search as a single complex relation. Overall, these annotation procedures allow for the full

spectrum of LVCs to be annotated appropriately, with the semantic arguments of the noun

accounted for, but the presence of the verb also noted, since it does contribute shades of meaning

to the construction.

2.3.3 PropBank LVC Annotation Schema: Annotation Guidelines

The preceding section gives an overview of the LVC annotation infrastructure, and this section

will now provide a description of the guidelines for LVC annotation that allow for consistent

recognition of LVCs (Bonial et al., 2013). The annotators are firstly provided with a detailed

theoretical description of LVCs, drawing upon existing linguistic descriptions of the

constructions. The annotators are also provided with this list of heuristics for recognizing LVCs:

i.   Does the noun object denote an event or state? If not, it is not an LV. If so, proceed to the next question.
ii.   Are the arguments of both the noun and the verb more representative of typical arguments of the verb relation or the noun relation? (e.g. *I made an offer to buy the house for $200,000* — such price arguments are more typical of *offering* events than *making* events). If the arguments are more typical of the verb relation, it is not an LV. If the arguments are more typical of the noun relation, proceed to next question. If you are unsure, here is a secondary test: try to add other common arguments from the verb's rolesets to see if they are semantically felicitous with the usage in question. For example, a Material argument is typical of *creating/making* events; thus, one could test the appropriateness of a Material argument in the usage: ?*I made an offer out of paper and ink*. If the addition of one of the verb's arguments does not seem appropriate, this indicates that the semantics are likely stemming from the noun and the usage is light. If the addition of the verb argument does seem plausible, this is evidence that the semantics are stemming from the verb and it should be treated as a heavy usage.

iii. Similarly, does the potential light verb link the eventive or stative noun to one of its typical semantic roles, realized by the subject of the verb? (e.g. *I* -Agent *made a call to the hospital*; *I* -Recipient *got a call from the hospital*). If the subject argument linked by the potential light verb seems to carry one of the semantic roles of the noun relation, it should be considered a light usage; proceed to the next question for a final check. If the subject argument does NOT carry one of the semantic roles typical of the noun relation, then this is evidence that its semantic role has been assigned by the verb, and the usage should be treated as heavy.

iv. Consider rephrasing the instance using a lexical verb related to the noun if one exists (e.g. *I called the hospital; I was called by the hospital*). If the rephrasing still captures the majority of the event semantics (even if there are small variations in voice, aspect or tone), then mark the instance as a light verb. Only if the verb adds semantics beyond what is captured by the eventive or stative noun should it be considered heavy. For example, *I got a call from him* should be considered light, while *I need a call from him* should not be annotated as an LVC. This test is helpful, but a counterpart lexical verb may not exist, so it is not a requirement.
(Bonial et al, 2013: 49-50).

If all or most of the characteristics outlined in these heuristics are met, the instance is considered a light usage. Notice that these guidelines already go far beyond the most typical test used in other annotation procedures, which is captured in (4): the rephrasing test. Further comparison of these guidelines to others will provided in the next section, Section 2.3.4.

The annotators are also provided with a list of LVC annotations for less typical LVCs that may be difficult for annotators to distinguish for the following reasons: 1) the eventive or stative noun may not have a clear verbal counterpart, so it is hard to rephrase with a lexical verb to get a sense of what the event really is, 2) the subject argument linked by the verb is not an Agent, and may be something as peripheral as a listener or Recipient, or 3) the light verb may seem to change the event's valency, similar to a passive voice construction. Despite these potentially misleading characteristics, usages like the following example, drawn from the PropBank guidelines, should be considered light:

29. I **had** no **right** to print that.

> ARG1: I
> ARGM-LVB: had
> ARGM-NEG: no
> REL: right
> ARG2: to print that
> (Bonial et al., 2013: 51)

Some of these confusing cases constitute borderline cases or "semi-light" verb usages, the treatment of which will be described and motivated in the next section.

Finally, the annotators are provided with a list of "tricky cases," involving what seem to be light verbs in abstract usages that are not truly light verbs, generally because the subject argument does not carry a thematic role licensed by the noun predicate, and instead carries an argument licensed by the verb. These include causal verbs that introduce a subject argument that is an outside causer. Two guideline examples are listed below with notes excerpted from the guidelines:

30. We've **had assassinations** before this.

> Note: Here, 'we' does not denote the person assassinated or the killer, but rather just the group of people affected by the assassinations. The best roleset for this case would be have.03, indicating abstract possession.

31. It **gives** you a **sense** of the terror she must have felt.

> Note: Here, 'give' introduces an outside causer that is not part of the eventive noun's roleset (something is causing you to sense something). Since a sense of what she felt is being metaphorically transferred or imparted, the best roleset is give.01. Similarly, *You gave me the ride of my life* introduces an outside causer of the 'ride' that is not part of the normal 'ride' roleset.
> (Bonial et al, 2013: 52-53)

Thus, annotators are familiarized with examples of borderline cases that are or are not LVCs and provided with a theoretically motivated explanation for why a usage does or does not qualify as

an LVC. Most of these tests for LVC-hood rely on the annotator considering the validity of

either verb or noun rolesets, and these are readily provided to the annotator for easy comparison

of whether the semantic roles assigned to arguments in a usage seem to be assigned by the verb

or the noun.

Although it remains difficult and subjective to determine how semantically neutral a verb

must be to act as a light verb, the success of these guidelines has been demonstrated in the high

agreement rates between annotators. On a task composed solely of the most likely light verbs

(*give, have, take, make, do*), agreement rates between annotators was 93.8%.[7] However, it

should be noted that most usages of these verbs are not light, potentially biasing agreement rates:

in a sample of 3,300 instances of *give, have, take, make, do*, 473 were identified as actual light

verbs, or 14%; other usages were either heavy senses or auxiliaries in the cases of *have* and *do*.

2.3.4 Comparison to Other Annotation Schemas

Because the precise definition of LVCs remains under debate, and black and white classification

of LVCs will always be difficult due to fuzzy boundaries, different annotation schemas have

arisen for the manual identification of LVCs. These varying schemas in turn result in different

inclusion and exclusion of particular usages, and differing data sets of LVCs that are potentially

used as training data for machine learning algorithms. In turn, this will result in distinct

successes and failures of automatic identification. The PropBank annotation schema and

resulting corpus of LVC annotations is compared here to two other existing schemas and data

---

[7] http://verbs.colorado.edu/propbank/ita/webtext-p25-SEL-LightVerb.html

sets: the TectoGrammatical Treebank (TTree) (Cinková et al., 2004) and the BNC data set

developed by Tu and Roth (2011).

The TTree annotation guidelines require a specific treatment for "Support verb

constructions," which are described as "multi-word predicates consisting of a semantically empty

verb which expresses the grammatical meanings in the sentence and a noun (frequently denoting

an event or a state), which carries the main lexical meaning of the entire predicate" (Cinková et

al., 2004: 91). The guidelines proceed with a description of the "rephrasing test" described in the

previous section, which indicates that most support verb constructions can be rephrased with a

one-word predicate. Although the guidelines point out that this is not a necessary condition, no

examples are given of support verb constructions that do not adhere to this condition. Next, the

logistical procedures for annotation are given. In a following section on "Quasi-control with

support verb constructions," a description of how to handle arguments of support verb

constructions makes it clear that under the TTree guidelines, the shared arguments of support

verb constructions can be licensed by either the verb or the noun predicate. This is quite distinct

from PropBank LVC guidelines, which define LVCs more narrowly as constructions wherein the

verb does not assign semantic roles to the arguments of the construction, but merely syntactically

licenses the subject argument.

This difference in views results in a very different data set of TTree Support Verb

Constructions from the PropBank LVC data set. In general, the TTree list of Support Verb

Constructions encompasses a much wider variety of support verbs beyond LVCs. A total of 694

support verb cases from TTree were therefore examined for this research, and refined into five

categories of support verb subtypes, including LVCs as defined by PropBank, and a category of

idiomatic expressions. This comparison provides for a nice overview of support verbs generally, and the distinguishing characteristics of light verbs specifically. Each of these categories of support verbs will be described in turn in the paragraphs to follow.

The first notable subcategory of support verbs is Semi-Light Verb Constructions (semi-LVCs). As outlined in the previous section, LVCs as defined by PropBank involve only arguments with semantic roles assigned by the noun predicate, and do not include cases where the verb contributes arguments that are not already intrinsic to the semantics of the noun. If the verb does contribute unique semantic roles, this is an indication that it retains semantic weight and should not be considered a light usage. In cases of semi-LVCs, the semantic roles associated with the verb and noun relation overlap, such that it is impossible to determine which is projecting the thematic roles. *Give* is a notable case that is often semi light. Consider for a moment the subcategorization frames of the dominant heavy sense of *give,* the transfer sense:

32. *I gave a book to her* - NP1[Agent] V NP2[Theme] PP[Recipient]

33. *I gave her a book* - NP1[Agent] V NP2[Recipient] NP3[Theme]

These are the typical arguments and alternations of this sense of *give.* If *give* is used in a LVC, then the subcategorization frame is quite different, with semantic roles stemming from the noun relation:

34. *I gave my blessing to them* -

    NP1[Agent] LV NP2[True Predicate] PP[Patient of True Predicate]

35. *I gave them my blessing* -

    NP1[Agent] LV NP2[Patient of True Predicate] NP3[True Predicate]

In this case, what would be the Recipient argument of the heavy usage is more accurately the

Patient of the true predicate; i.e. *them* is more accurately interpreted as the *thing blessed,* as

opposed to the Recipient of an abstract item that is a *blessing*. This is not to say that the two

interpretations are not related, or that the latter interpretation is "wrong," but for the purposes of

Natural Language Understanding, recognizing a *blessing* event provides much more semantic

specificity and allows for more appropriate inferences than interpreting these as *transfer* events

with an abstract Theme. If *give* is used in a semi-LVC, both *give* and the predicating noun share

the same typical semantic roles, and it is unclear whether the semantics stem primarily from the

noun or the verb:

36. *I gave a hint to them.* - NP1[Agent] LV NP2[Noun Predicate/Abstract Theme] PP[Recipient]
37. *I gave them a hint.* - NP1[Agent] LV NP2[Recipient] NP3[Noun Predicate/Abstract Theme]

In this case, unlike that of the true LVC above, *them* can be understood as the Recipient of either

give or *hint* (i.e. *give to them/hint to them*)*,* but cannot be understood as the Patient or Theme of

hint — them is not the *thing hinted*. Similarly, the Agent subject could be assigned by either

relation.

      PropBank has, in the past, been relatively conservative in its definition of LVCs, and

since these semi-LVCs fit the semantics and syntax of heavy usages, these usages have been

annotated using the heavy verb sense (in this case, the transfer sense), and then the eventive noun

is annotated in a separate pass. Although this does capture most of the semantics of each

predicate, non-LVC verb and noun annotation is restricted to the domain of locality of either the

noun or the verb. Thus, in semi-LVC cases like this, the *hint* annotation would not include the

Agent argument, which is within the domain of locality of the verb, but outside the domain of

locality of the noun. This not only leads to some gaps in the semantic representation, but also

some inconsistencies in annotation across light verbs. For example, *make a speech* is treated as

an LVC (since the arguments of *make* and *speech* do not overlap), but *give a speech* is treated as

a transfer event of an abstract object (because the arguments do overlap). Having noted this

inconsistency and considered the ramifications for PropBank serving as training data for

recognizing LVCs, we will loosen the guidelines for LVC annotation and revisions will be made

to annotate these usages in the same fashion as LVCs — recognizing the combination as a

complex predicate.

The second subcategory of support verbs noted in the TTree annotations was that of

aspectual verbs. These verbs carry clear aspectual meaning that adds notable semantic content to

the construction. Of course, all verbs have inherent aspect, but these verbs have aspectual

meanings in the sense that they carry the semantics of beginning, continuing, or ending another

event or state. For example, the TTree cases include the construction *hold out hope*, as in:

37. Right up till the 11th hour I **held out hope** that Google would give Reader a last minute stay

     of execution.

The verb phrase *hold out* carries the semantics of continuation that would not be included in an

LVC, such as *I had hope…* As a result, in PropBank, these verbs should be annotated with an

aspectual sense/roleset that captures their semantics of continuation. However, like the semi-

LVCs, this entails that the eventive or stative noun will be annotated for arguments only within

its domain of locality, which again will exclude the Agent "I" from constructions like this. This

suggests that perhaps PropBank should consider a more general treatment of support verbs,

wherein even verbs with some semantic content can serve as syntactic links to other arguments

of a predicating noun.

A third category of TTree support verbs is copular verbs, specifically the verb *be* in the

TTree annotations (other copular verbs would be *become, seem,* and certain senses of *look, smell,*

taste, etc.). For example, TTree lists *be right* as a support verb construction, as in:

38. Tell me if **I'm right** about this, New York Times…

These cases are often notably distinct from the PropBank definition of LVCs in that the

complement is a predicate adjective rather than an eventive or stative noun. Nonetheless, these

are certainly cases where PropBank annotation should recognize the lack of semantic content

contributed by the verb, and instead capture the semantic content contributed by the predicate

adjective. Previously, PropBank simply annotated the copular verb with a copula sense, and

annotated the Topic (subject) and Comment (complement) respectively, without decomposing the

semantics of the adjective. More recently, PropBank has expanded into the annotation of

adjective relations, thus each annotation of a copular verb has undergone a second pass of

annotation, similar to that of the second pass of annotation for LVCs. In this pass, like LVCs, the

syntactic arguments of both the adjective and the support verb are annotated with a roleset

associated with the adjective. This effectively recognizes the supporting role of the verb and the

primary semantic contribution of the adjective, but also recognizes these as a distinct

construction from PropBank LVCs.

Verbs that add outside causation make up a fourth category of support verbs detected in

the TTree annotations. Although quite similar to light verbs, these verbs assign a Cause role to

the subject that is not intrinsic to the semantics of the eventive or stative noun. For example,

grant right is found in the TTRee annotations, as in:

39. This would **grant** them the **right** to engage in sympathy strikes.

Notably, the roleset for *right* does not include an outside Cause enabling that right:

> **Roleset id**: right.01, *something due to you by law, tradition or moral principle*
> **Roles**:
> Arg1-PPT: entity with a right
> Arg2-GOL: what they have a right to

Therefore, while *have a right* would be an LVC, the Cause subject argument linked by *grant*

indicates that this verb carries some semantic weight. Thus, this is another case where the verb

should be annotated with a roleset that captures the semantic roles it carries.

Finally, it should be noted that some of the TTree annotations of support verb

constructions are likely better treated as idiomatic expressions. These combinations lean towards

a purely non-compositional meaning, evidenced by the fact that it would be strange to create a

roleset for any of the elements that could be considered the true predicate in a support verb

construction. One example from TTree is *give shoulder,* as in:

40. He gave me the cold shoulder the whole evening at the party.

It seems inaccurate to attribute the semantics of *ignore* to any of the elements of this

construction: *give, cold,* or *shoulder*. Thus, in these cases, PropBank would mark *give* with a

roleset available to all verbs that indicates it is part of a non-compositional Idiomatic Expression:

IE. We are currently exploring the possibility of capturing the semantics of Idiomatic

Expressions by "aliasing" such expressions to existing verbal rolesets that capture a similar

meaning, in this case, possibly Ignore.01 (Bonial et al., 2014b).

In summary, the TTRee support verb annotations include a much wider variety of support verbs, with varying levels of semantics that are attributable to the verb. Nonetheless, TTree benefits from a very consistent treatment of all cases where verbs syntactically support potential arguments of a predicating noun. PropBank will certainly be adopting a similar treatment for the semi-LVC cases, and should consider modifying its treatment of aspectual and causal verbs to recognize both the semantic contribution of the verb and its syntactic support of an argument of the nominal complement.

A second annotation schema that PropBank should be compared to is that of Tu and Roth (2011), as this data set has been used to establish the state of the art LVC detection system (86.3% F-score), and this is one goal of the PropBank annotations. As mentioned briefly in Section 2.2.3, Tu and Roth construct a dataset of 2,162 English sentences with LVCs drawn from the British National Corpus (BNC). Their approach in constructing this data set differs from that of PropBank in several ways, and therefore results in resources containing some overlapping constructions and some constructions that are unique to each resource. Firstly, the authors restrict their data set to LVCs involving the 6 most frequent light verbs: *do, get, give, have, make, take*. In the PropBank annotation process, it is possible for annotators to mark any verb as a light verb, resulting in a data set that contains LVCs with a greater variety of verbs, including expressions like *bring charges against* and *conduct repairs*. Therefore, the PropBank data contains 7 LVC types with verbs not included in the Tu and Roth data. Secondly, Tu and Roth filter their data set by including only LVCs with nouns that are zero-derived nominals (the same word can be used either as a noun or a verb, e.g. *offer*), or nouns that are derivationally related to a verb (e.g. *destruction*). The PropBank data set again includes a bit

more variety, since LVCs involving nouns that have no etymologically related verb counterpart are included, such as *take a trip*. Thus, the PropBank data includes 25 LVC types with nouns that would not be found in the Tu and Roth data.

Although the PropBank procedure allows for more variety, this has not resulted in a broader data set with more unique LVC types overall. The comparison shows that there are 115 LVC types that appear in both data sets, 245 LVC types that appear only in the BNC, and 218 LVC types that appear only in PropBank. Although some of these different types may simply arise from the differing sources and genres, there are notably more instances of LVCs involving *get* and *give* in the BNC data set. As discussed in the TTree comparison, PropBank has previously treated many usages of both *get* and *give* as semi-LVCs, and therefore opted to annotate both the argument structure of the verb and that of the noun in distinct annotation passes instead of marking these as LVCs. Thus, there is a clear difference between the data sets in the number of LVC types involving *get* and *give*: the BNC data set includes 83 additional types with these verbs. Because the PropBank practice has also led to inconsistencies in the treatment of similar constructions across light verbs, it should be noted again that PropBank will be loosening the annotation requirements and including such semi-light usages in the LVC annotations. In summary, the Tu and Roth data and PropBank differ in ways that likely provide an advantage for the Tu and Roth detection system: the BNC data has less variety in the range of light verbs and nouns that can be involved in LVCs, and it embraces a more general definition of LVCs since it does not distinguish light and semi-light usages.

2.4 LVC Detection System Trained on PropBank

The true proof of the quality of PropBank's LVC annotations would be the successful training of

an automatic LVC detection system (Chen, Bonial & Palmer, submitted). In joint work with a

computer science PhD student, Wei-Te Chen, this system has been developed and refined to the

point where it achieves an F-Score of 89%, superior to the state of the art system developed by

Tu and Roth (2011), who achieve an F-Score of 86.3%. My role has been to assist in the

selection of features, and to aid in the use of linguistic resources for developing these features.

The linguistic resources used in the development of features, the features themselves and

performance of the system are described in the sections to follow.

2.4.1 Linguistic Resources

This system makes use of several existing corpora: PropBank, which has already been described

in detail, the OntoNotes sense groupings (Pradhan et al. 2007), WordNet (Fellbaum, 1998) and

the British National Corpus (BNC). A bit of background information is given on WordNet and

the OntoNotes sense groupings in the paragraphs to follow.

WordNet (WordNet) is a large electronic database of English words,[8] which was, in part,

inspired by work in psycholinguistics investigating how and what type of information is stored in

the human mental lexicon (Miller 1995). WordNet is therefore an attempt to create a searchable

electronic lexicon that is organized according to relations that are hypothesized to also organize

the mental lexicon. WordNet is divided firstly into syntactic categories: nouns, verbs, adjectives

and adverbs, and secondly by semantic relations. The semantic relations that organize WordNet

are: synonymy (given in the form of 'synsets'), antonymy, hyponymy (e.g. a Maple is a

---

[8] http://wordnet.princeton.edu/wordnet/

tree; therefore, tree is a hypernym of Maple), and meronymy (part-whole relations). These

relations make up a complex network of associations that is both useful for computational

linguistics and NLP, and also informative in situating a word's meaning with respect to others.

Of particular interest for this research are the synsets, the hyponymic relations of nouns in

WordNet, and the noun's 'type,' as indicated by the lexical file information: for each noun in

WordNet, lexicographers have coded the noun with one primary superordinate, or lexical file,

given forty-five numbered options. WordNet is then generated by taking this relational

information as a starting point. In our research, nouns that can possibly denote events or states

are the focus, because it is these nouns that can theoretically combine with a light verb to form

an LVC. The types hypothesized to designate eventive and stative nouns, used in this research,

are listed in Table 2.4.1. The manners in which the synset, hyponym and lexical file information

(or noun 'type,' as it will be referred to throughout the rest of this document) are used will be

described in the next section, Section 2.4.2.

Lexical File Noun Type	Nouns denoting…
noun.act	acts or actions
noun.cognition	cognitive processes
noun.communication	communicative processes
noun.event	natural events
noun.feeling	feelings and emotions
noun.location	spatial position
noun.motive	goals
noun.phenomenon	natural phenomena
noun.possession	possession and transfer of possession
noun.process	natural processes

Lexical File Noun Type	Nouns denoting…
noun.relation	relations between things
noun.state	stable states of affairs

Table 2.4.1: WordNet noun types hypothesized to correspond to potentially eventive or stative nouns that may be involved in LVCs.

The OntoNotes corpus integrates several layers of different annotation types in a single corpus, making it ideal training data for semantic analysis (Pradhan et al. 2007). The five layers of annotation include: 1) the syntactic parse from the Penn Treebank (Marcus, Santorini, and Marcinkiewicz, 1993), 2) proposition structure from PropBank, 3) coarse grained word senses from the OntoNotes sense grouping inventory, 4) named entity types, and 5) anaphoric coreference. The latest release, the OntoNotes 4.99 corpus (Weischedel et al. 2011), contains 2.6 million English words. In this research, the PropBank and word sense layers are of primary interest, the latter is described next.

The OntoNotes sense groupings can be thought of as a more coarse-grained view of WordNet senses. This is because these sense groupings were based on WordNet senses that were successively merged into more coarse-grained senses based on the results of inter-annotator agreement in tagging of the senses (Duffield et al. 2007; Pradhan et al. 2007). Essentially, where two annotators were consistently able to distinguish between two senses, the distinction was kept. Where annotators were not able to consistently distinguish between two senses, the senses were reorganized and tagged again. Each OntoNotes sense lists which WordNet senses it includes. This sense inventory was used to annotate verbs with more than three senses in WordNet. It was found that sense distinctions with this level of granularity can be detected automatically at 87-89% accuracy, making them effective for NLP applications (Dligach and

Palmer 2011). Unfortunately, the sense tagging is not complete for all of the OntoNotes corpus: there are about one million verbs and nouns in OntoNotes 4.99, but only 288,217 of these have sense tags (although many are surely monosemous), including 120,400 nouns with sense tags.

2.4.2 Features

Three different types of features are used by the classifier: Basic Features, OntoNotes Word Sense Features, and WordNet Features. Basic Features include the lexicon, part of speech (POS) tags, the dependency relation of a potential light verb and an eventive or stative noun, as well as the paths of the dependency relations, and the subcategorization frame, which concatenates the dependency labels of the light verb and noun. OntoNotes Word Sense features allow for distinct senses of nouns to be recognized, which is very important for distinguishing surface-identical constructions that may or may not be LVCs. For example:

41. We are going to **take** a **look** at the trials and tribulations of Martha Stewart.

42. Barbie gets a makeover to **give** her a more youthful **look**.

The first sentence is an LVC, while the second is not. The difference in LVC status is reflected in the two different senses of *look*: the meaning of the first *look* is "act of looking," and the second usage is closer to the meaning "perceived appearance." Despite identical surface forms, these usages are distinguished by OntoNotes word sense tags. For some experiments with the classifier, the test data used automatically generated dependency trees that don't have any sense tags. For these cases, the automatic Word Sense Disambiguation model of Lee (2002) was adopted and applied, performing with an F-score of 73.66%, which is not ideal, but is still helpful since the word sense feature is only one feature supporting the improvement of LVC

recognition. WordNet features include WordNet sense information, which is extracted via

mappings between OntoNotes word senses and WordNet senses. The WordNet noun type (lexical

file information) was also used, as this allows for a focus on eventive and stative noun types that

can potentially serve as complements in LVCs. Additionally, WordNet hyponymy was used —

the hypernym of each noun provides a more generalized feature than the WordNet sense itself,

but more fine-grained information than the base noun type.

2.4.3 Classifier Performance: Results

In order to best compare our results to the state-of-the-art standard set by Tu and Roth (2011) our

model was first trained and evaluated on the same BNC data used by Tu and Roth (which

includes 1,039 positive LVC examples and 1,123 negative LVC examples). Table 2.4.3-1 gives

the results of our classifier on the BNC data set at each step: precision, recall and F-measure.

The baseline model involves basic features only. Our *All features* model, which includes the

three WordNet features, gains around 3-4% improvement for classification of positive and

negative examples, in comparison to Tu and Roth's models with contextual features and

statistical features. The best results are shown in bold-face: our F-score of 89% outperforms the

best Tu & Roth results.

Model	Positive/Negative LVC Identification	Precision	Recall	F-Measure
Tu & Roth - Contextual	+	**86.49**	84.21	85.33
	-	86.15	**88.19**	87.16
Tu & Roth - Statistical	+	86.48	85.09	86.46
	-	86.72	87.40	87.06
Basic	+	81.13	86.00	83.50
	-	88.89	84.85	86.82
All Features	+	85.32	**93.00**	**89.00**
	-	**94.31**	87.88	**90.98**

Table 2.4.3-1: Classifier results on Tu & Roth's BNC data set, compared with both the contextual and statistical classifiers developed by Tu & Roth.

Our system was then trained and tested on the OntoNotes corpus (which includes 1,768 positive LVC examples — 1,588 used for training and 180 used for testing), with and without gold standard dependency trees. With automatic parse trees, we achieved Precision of 54.94%, Recall of 77.22% and an F-Score of 64.20%. These results are lower, in part, due to errors in the automatic parses. With gold standard dependency trees, we achieve an F-Score of 80.68%. Clearly, even with gold standard dependency trees, these results are lower than those obtained on the BNC data set because the OntoNotes data is both more challenging and more representative of real language, as it is not a balanced data set, and it includes a wider variety of LVCs as discussed previously. Table 2.4.3-2 gives the incremental feature contribution for the system that was trained and tested on the gold standard dependency trees.

Feature	Precision	Recall	F-Measure	Diff (%)
Basic	78.09	78.09	78.09	-
WordNet-Sense	80.23	**79.78**	80.00	+1.91
WordNet-Type	80.68	**79.78**	80.23	+0.23
WordNet-Hyper	81.61	**79.78**	**80.68**	+0.45
OntoNotes Word Sense	**81.77**	78.09	79.89	-0.79

Table 2.4.3-2: Incremental feature contribution for the system trained and tested on the OntoNotes data set with gold standard dependency parses.

All three WordNet features contribute to the F-Score incrementally, and after these are added, the model reaches the best F-Score of 80.682. Although the addition of the OntoNotes Word Sense feature decreases the F-Score, it increases precision.

2.4.4 Using WordNet Relations to Find Previously Unattested LVCs

Although the above model could capture the majority of the LVCs in the corpus, those that are detected are relatively high-frequency LVCs. This led to the question of whether or not there is a better way to detect previously unattested LVCs. In the previous section, the results show that WordNet features provide positive contributions to our model. In this section, we analyze a small set of data from the OntoNotes corpus and corresponding WordNet features to explore the following possibility: if there is a high-frequency, attested light *verb + noun* combination, then any other eventive or stative noun sharing a synset or hypernym with this noun may also combine with that light verb to form an LVC. To explore this, we first calculate the frequency of all the gold LVC pairs in the OntoNotes 4.99 data. Then we extract the top 10 highest-frequency occurrences of *light verb + noun* pairs. In order to generate candidate LVC pairs, we fix the verb

found in the high-frequency, attested LVC, and combine this with nouns that either share a synset

or a hypernym with the noun from the same high-frequency LVC. This replacement of the

eventive noun with its synonyms could allow for the discovery of promising LVC candidates.

For example, the derived LVCs *make attempt, make effort,* and *make endeavor* are obtained by

examining the synset and hypernym relations of the high-frequency LVC *make contribution*.

Using this process, we obtain a resulting total of 91 tokens of potential LVCs in the OntoNotes

corpus. When we compare these to our existing annotations, we see that 49 of these are already

annotated as LVCs. Table 2.4.4 displays the numbers of gold, true LVCs and candidate *verb +*

noun pairs. The results show that the ratio of LVCs generated from the synonyms of high-

frequency LVCs is twice the baseline probability that any *light verb + noun* pair is a LVC. Thus,

we can assume that WordNet synsets could play an important role in discovering low-frequency

and previously unattested LVCs.

Data	# of *Verb + Noun* Pairs	# identified as LVCs	Ratio (%)
Light Verb + WordNet Synonym	91	49	53.85
Light Verb + Any Noun	8,191	1,911	23.31

Table 2.4.4: *Verb + noun* combinations derived from replacing existing LVC nouns with WordNet synonyms, listed with the number that were previously identified as LVC tokens. This is compared with the number of all combinations of previously recognized light verbs with any noun, listed with the number of these cases that were previously identified as LVC tokens. Notably, the ratio of actual LVCs to other combinations is much higher when the WordNet information is used.

Although promising, it should be noted that of the 91 potential LVC tokens that this

process generated, there were only 22 unique *verb + noun* types. Of these 22 potential LVC

types, only 4 types were attested in the corpus, and already annotated as LVCs. These include the

relatively high-frequency types *make effort* and *make commitment*, which account for the

majority of the 49 existing LVC tokens detected in this fashion. This is not to say that the other

candidate LVC combinations created via this process are not LVCs, but they are either not

attested or not annotated as LVCs in the corpus. Further research is required. In Chapter 4, the

question of LVC productivity is examined in greater detail, along with the acceptability of LVCs

formed in a similar fashion to what was done here — by holding the verb constant and replacing

the verb with a synonym — however, synonyms are detected using FrameNet instead of

WordNet.

Chapter 3
LVC Function (Research Question 2)

We have noted that cross-linguistically, it is common for LVCs to exist in complementary

distribution with lexical verbs: LVCs are exploited only where a lexical verb counterpart to the

eventive noun does not exist (Zarco, 1999). In languages where this is the case, the semantic

function of LVCs is fairly clear: as the meaning of the LVC is quite distinct from any lexical

verb, the LVCs fill an expressive void in the language's resources. However, in English, lexical

verbs and LVCs exist where the difference in meaning is much more nuanced and arguably

difficult to pinpoint. Consider the following examples, drawn from the Corpus of Contemporary

American English (COCA) (Davies, 2008):

43. For heuristics, researchers have **investigated** a number of ideas using machine learning...
44. ...This inspector general **did** a rather controversial **investigation** of a friend and political supporter of the president...

45. 12 hours into the siege, darkness was beginning to set in, and police **decided** to make a move.
46. The governor **made** a **decision** to release his tax records.

47. She **appeared** with me on VH1 "Celebrity Rehab."
48. This afternoon, Bahrains King Hamad **made** a rare **appearance** on television.

49. And so I **walked** over to the end of the driveway with him...
50. At 5 AM, we **took** a **walk** through the parking lot.

51. Most people **laughed** about it.

52. Maureen, we've all **had** a good **laugh** over it at their expense…

53. ...He **spoke** of the importance of having a choice of health insurance companies.

54. Vice President Biden **gave** a **speech** on foreign policy at NYU this morning.

55. Silas **groaned** as he extricated himself from the vehicle.

56. Lyra **gave** an irritated little **groan** and buried her face in her pillow.

These examples demonstrate that while the semantic difference between the LVC and lexical verb counterparts may vary in extent and quality, for the most part it seems that the speakers and authors of these utterances could have expressed very similar concepts using either the LVC or lexical verb option. Of course, the truth of the previous statement depends upon the nuances of meaning expressed in the LVC as opposed to the lexical verb, the subject to be explored next.

Presumably, if two types of predication, or two distinguishable coding means generally, exist in a language, then they are not identical in meaning (Grice, 1975; Cattel, 1984; Goldberg, 1995; Frajzyngier & Shay, 2003). Given this assumption, the question remains of why these two competing forms remain in the language, especially when the lexical verb is arguably a more parsimonious, efficient form of expressing the same concept. According to both the Gricean Maxim of Quantity (Grice: 1975:45-46) and Zipf's notion of Speaker's Economy (1949:20), one should make a conversational contribution that is as informative as required, and not more informative than required. Similarly, according to the Gricean Maxims of Manner, one should be brief (1975: 45-46). To some extent, the use of LVCs when the lexical verb could be used violates these conversational maxims. This chapter employs a corpus study to examine how the

usages of LVCs and counterpart lexical verbs differ, which may motivate the use of an LVC over

a lexical verb. Before turning to the corpus study, existing theories on the function of LVCs in

comparison to their verbal counterparts are discussed. These break down into two distinct, but

not opposing trajectories: many synchronic linguistic studies point to aspectual differences

between LVCs and verbal counterparts, while many diachronic studies indicate that LVCs seem

to be easier to modify than lexical verb counterparts.

3.1 Aspectual function of Light Verb Constructions

One of the most widely proposed semantic differences between lexical verbs and LVC

counterparts are aspectual differences: most commonly, the LVC counterpart is thought to lend a

sense of boundedness, or telicity to the event denoted. Butt (2003) points out that the LVCs

crosslinguistically tend to affect the Aktionsart of the joint predication by rendering the event

bounded. Similarly, according to Wierzbicka, the clear difference between the lexical verbs and

the periphrastic *have a V* construction is that the latter presents the action or process as limited in

time (1982: 757). Several authors postulate this difference as the key motivation for LVCs as a

distinct mode of expression: "...the characteristic [of LVCs] which is most consistent and which

perhaps makes for its widespread use, is aspectual connotation" (Live 1973: 34). Brinton &

Akimoto (1999) cite LVCs in English as an important means of making situations telic,

specifically by converting activities into accomplishments or achievements, without the necessity

of stating an explicit goal. In addition to adding telicity (Prince 1972), perfectivity is also

occasionally associated with the LVC in contrast to the lexical verb (Stein, 1991; Prince 1972).

In this research, whether or not an LVC or counterpart lexical verb was telic, with an inherent endpoint, was considered through the lens of research on Aktionsart from VanValin and LaPolla (1997). To determine whether or not a given usage of a predicate was telic, tests for determining Aktionsart type were employed (1997: 94). Both achievements and accomplishments are [+telic], a feature indicating that the predicate depicts a state of affairs with an inherent terminal point (1997:93). VanValin and LaPolla point to two specific tests that distinguish telic from non-telic predicates:

Occurs with *X for an hour, spend an hour Xing.*
Occurs with *X in an hour.*

These tests isolate the properties of duration in time and whether or not an event has terminal endpoints. VanValin and LaPolla (1997) indicate that while states and activities readily take the *for*-phrase duration, telic activities and accomplishments are generally only compatible with the *in*-phrase duration, and sometimes only very short durations (e.g. *in the blink of an eye*). Thus, if an LVC or counterpart verb usage was compatible with the first test, but not compatible with the second test, it was determined to be [-telic] (e.g. *I thought about aspect for an hour,* vs. ?*I had a thought about aspect for an hour*). There is a certain amount of aspectual indeterminacy involved in some LVCs. For example, it does not seem incorrect to say *I took a shower for twenty minutes* or *I took a shower in twenty minutes.* For this reason, this research does not attempt a full empirical analysis of the aspectual properties of LVCs, instead focusing on broader descriptions providing evidence for and against the possibility of an aspectual function of LVCs.

Aspectual differences are clearly found in certain LVCs of English. Consider the following examples of the activity verb, *think,* and its counterpart LVC *have a thought* (drawn from COCA):

57. I **thought** that the advice I was giving my clients was special. (-telic)
58. Melinda **had** a **thought**: Maybe there would be some way for her husband to collect Mann's DNA... (+telic)

While (57) gives the impression of an event that may have continued for an undetermined time, (58) gives the impression of an endpoint of the moment when the "thought" arises. However, if we consider the very same LVC (*have + thought*), we can see other examples where the construction does not seem to lend telicity:

59. She also **had thoughts** of suicide. (-telic)
60. ...She **had** violent **thoughts** about the woman she saw as breaking up her family. (-telic)
61. I was just **having** really, really bad **thoughts** towards the enemy. (-telic)

In these cases, where "thought" is plural and lacks a determiner, the LVC doesn't necessarily give an impression of a bounded, telic event, nor does it necessarily carry the meaning of iterative, bounded periods of thinking. In the final example above, the light verb is in the progressive, indicating that the event is durative rather than punctual (Van Valin & LaPolla, 1997).

If LVCs exist as a resource in the English language separate from lexical verbs because LVCs lend telicity to the event, then the assumption follows that we would see a consistent distinction in telicity between the lexical verb (-telic) and LVC (+telic). As demonstrated above, this does not seem to be true for all realizations of a single LVC, and it certainly does not seem to be true of all LVCs. Clear counter-evidence comes from LVCs characterized by nouns that denote states (*understanding,* see 62-63 below, from COCA) or events that are already inherently telic (*decision, appearance,* see examples 64-67, from COCA):

62. He is pro-life and he **understands** business policies. (-telic)
63. President Karzai **has** a very good **understanding** of his own country. (-telic)

64. 12 hours into the siege, darkness was beginning to set in, and police **decided** to make a
    move. (+telic)
65. The governor **made** a **decision** to release his tax records. (+telic)

66. She **appeared** with me on VH1 "Celebrity Rehab." (+telic)
67. This afternoon, Bahrains King Hamad **made** a rare **appearance** on television. (+telic)

Thus, it seems as though the facts of English do not point to a consistent transformation from an atelic lexical verb to a telic LVC counterpart.

The presence or absence of articles accompanying the noun seems to play a role in the interpretation of telicity. Indeed, the outward evidence of the difference in telicity that is often pointed to is the presence of articles modifying the eventive/stative noun, which is taken to indicate the countable, bounded nature of the situation expressed (Brinton & Akimoto, 1999: 6;

Wierzbicka, 1982; Butt & Geuder, 2001). However, a an examination of article use with LVCs

shows that articles are not used in a consistent manner within LVCs, as we would expect if the

primary function of LVCs was to lend boundedness to the event or state described. Some LVCs

almost always occur with articles, some optionally occur with articles, and others never appear

with articles. Although some behaviors seem idiosyncratic, others seem to relate primarily to the

nature of the noun (mass or count) and whether or not the noun is plural:

68. He made a decision.
69. *He made decision.
70. He made decisions.

71. He made progress.
72. *He made a progress.
73. He made some progress.

Even in cases where the noun involved is an unbounded mass noun, the LVC encompassing this

noun is not necessarily telic, again indicating that there is no consistent relationship between the

aspectual quality of the event or state denoted by the noun and the resulting aspect of the LVC.

For example:

74. She takes advantage of the situation. (-telic)
75. She takes steps to change the situation. (-telic)

Thus, given the lack of consistency in the nature of articles within LVCs, it seems somewhat

unlikely that we can point to articles as evidence of an aspectual function of LVCs.

Additionally, if articles played an important role in the telic nature of LVCs, then one

may expect article use within LVCs to have been constant since the introduction of LVCs into the

English language. However, the diachronic picture shows that article use in LVCs seems to be

more closely related to the origin of the phrase than any aspectual considerations. Matsumoto

(1999) points out that in Middle English (ME), the use of articles in the language was unsettled.

In ME, Matsumoto argues, indefinite articles are sometimes dropped, and this doesn't appear to

change the meaning of an LVC (as seen in glosses for the same phrase with or without articles),

for example:

76. maken (a) noise

    'make (a) noise'

    (Matsumoto, 1999: 88)

This construction is found originally in French without the indefinite article (Mustanoja, 1960:

271, cited in Matsumoto,1999), and variants with and without the indefinite article appear in the

works of Chaucer and Malory. Matsumoto writes,

> "Even for a Modern English [Complex Predicate (CP)], it is difficult to judge whether the
> CP expression involves aspect or not, much less for an ME CP, because the system of
> articles had not fully developed at this stage. I think that in ME the presence or absence
> of articles in CPs is a problem of articles, not CPs" (1999: 88).

The "problem" of articles is further complicated by factors of grammaticalization and idiomaticization throughout the history of English. This warrants a diachronic overview of LVCs, to follow.

3.2 Diachronic Considerations

Although the number of LVCs increased dramatically in the Modern English (ModE) period (Mathesius, 1975: 1040), similar expressions did exist in Old English (OE). For example, the following LVCs are attested in the late OE period:

77.   answare sellan
         answer   give
          'give answer'
78.   reste habban
         rest   have
          'have rest'
79.   ware niman
         care take
          'take care'
(Akimoto & Brinton, 1999: 23)

Like the LVCs of English today, the light verbs of OE also had "heavy" senses in which they predicate fully, and the LVCs generally had a very similar lexical verb counterpart. Akimoto & Brinton find that 77 of 114 LVC types have "a relation of synonymity" between the LVC and corresponding lexical verb (1999: 50). Furthermore, in almost all cases where there is a

detectable difference in meaning, this difference was one of transitivity, often where the LVC did

not require an object (e.g. *give invocation* vs. *invoke X*) (1999: 50). The authors also note that

approximately one third (24 of 64) of the attested nominal complements occurred in combination

with more than one light verb, and the choice of light verb seems to have little effect on the

meaning of the LVC. In fact, Akimoto & Brinton (1999) indicate that the same complements

with light verbs of opposite meaning (*sellan* 'give' and *niman* 'take') were used equivalently in

the past:

80. and he heom **gryð sealde**

    and he them truce gave

    'and he made a truce with them'

81. **name gryð** and fryð wið hy

    take   truce and peace with them

    'truce and peace were made with them'

(Akimoto & Brinton, 1999: 48, 52)

Akimoto & Brinton do cite the importance that modification seems to play in the preference for

LVCs: noting that most LVCs are characterized by adjectival modification of the noun,

relativization, or conjoined nouns. Given this fact, the authors consider ease of modification and

facilitation of conjunction the primary motivation for use of LVCs in OE (1999: 51-2).

      In Middle English (ME), LVCs become much more frequent, although the light verbs

themselves had shifted somewhat due to rivalries between lexical items for *take* (*niman, tacan*)

and *give* (*sellan, yeven*) (Akimoto & Brinton, 1999). The most common light verbs observed in

ME are: *haven* 'have,' *taken* 'take,' *maken* 'make,' *don* 'do,' *yeven* 'give' (Matsumoto, 1999).

Compared with the 114 LVC types observed by Akimoto & Brinton (1999) in OE, Matsumoto

(1999) finds that there are 990 LVC types observed in the texts of ME, with many stemming

from the works of Chaucer and Malory. Matsumoto observes that LVCs seem to be a preferred

expression in the works of Chaucer and Malory, and believes that LVCs are used as a poetic,

stylistic device (1999: 61). This seems fitting given certain metrical considerations involved in

such literary works. In ME, there remain lexical verb equivalents to LVCs, and Matsumoto

specifically compares LVCs with adjectival modifiers to equivalent lexical verbs with adverbial

modification, and concludes that…

> "...Authors preferred, for stylistic reasons, a modified CP expression to a simple verb
> modified by an adverb. A modified CP rather than a modified simple verb was probably
> rhythmically easier even for speakers in Medieval times to produce, just as a modified CP
> is less difficult and awkward than a modified simple verb in ModE" (1999: 85-86).

In fact, constructions with lexical verbs along with adverbs were extremely rare in Chaucer and

Malory. However, Matsumoto theorizes that for LVCs that seemed more idiomaticized through

frequent use, the status of the LVC becomes more similar to that of a single, unified verb, and the

tighter connection of *light verb* + *noun* allows for the LVC to take additional adverbial

modifiers, as opposed to adjectival modifiers of the noun. Consider the following usage of the

frequently used LVC *make vow*:

82.   **make** myne **avowe** deuotly to Criste

      make my   vow   devoutly to Christ

      'make my vow devoutly to Christ'

       (Matsumoto, 1999: 87)

Thus, Matsumoto demonstrates that while adjectival modification seems to be an important factor in the preference for LVCs, the presence or absence of adjectival modification and adverbial modification may be affected by how frequent and fixed, or idiomaticized, the expression is.

Article use in ME, as mentioned previously, seemed to be optional, and seemed to play no role in the meaning or interpretation of the LVC (Matsumoto, 1999: 88). The expansion of the use of LVCs in ME can be attributed in part to French influence on English, as many LVCs seemed to be direct calques from French LVCs involving *faire* 'do, make.' Indeed, throughout ME, the light verb *maken* was the most frequent and collocated with the greatest variety of nouns denoting events or states; furthermore, the majority of noun collocates with the light verb *make* in ME were nouns of French origin (Tanabe, 1999: 106). Across all LVCs of ME, the presence or absence of an article was generally consistent with the French original; therefore, Tanabe concludes that one cannot ascribe the lack of an article to an idiomaticization process in which the noun is decategorialized and loses its characteristic determiner (1999: 118; Hopper, 1991). Rather, the presence or absence of an article seems to have a stronger relationship with the meaning or nature of the noun, a property consistent across French and English. Tanabe notes that nouns denoting continuing activity or mental attitudes generally do not take an article (e.g. *have knowleche of*, *have consideracion of*) (1999: 119).

In early Modern English, the number of new LVCs entering the language begins to decline as English begins to evolve new phrases based on existing patterns (Prins, 1952). In Hiltunen's survey of LVCs through early Modern English, he concludes that LVCs without

articles are the oldest, based on both OE and French LVCs lacking articles (1999: 162).

Prins' (1952) survey of the phrases entering English from French does show that while a

minority of French LVCs involved possessive determiners (e.g. *faire ses dévotions,* 'do ones

devotions'), only four (of 96 total LVCs examined in Prins (1952), distinguished from non-LVC

phrase types by the present author) French LVCs entering the English language were optionally

characterized by an indefinite article (e.g. *faire une complainte,* 'make a complaint'). As the

specifying function of the article evolved and was normalized in the English language, LVCs

with indefinite articles were likely introduced (Hiltunen, 1999).

Overall, the diachronic picture demonstrates that article use within LVCs is inconsistent

in OE and ME, and the presence or absence of articles seems to be most strongly associated with

the presence or absence of articles in the French original, which reflects certain semantic

properties of the noun. In the French originals, we see a strong preference for LVCs without an

article, indicating that articles likely do not play an aspectual function in French LVCs.

However, a common factor of LVCs throughout the history of English is the presence of

adjectival modification. While all authors cited herein noted this frequent characteristic of

LVCs, Kyto (1999) is one of the few to indicate an actual percentage of modified LVCs in Early

Modern English. Kyto finds that overall, approximately 70% of LVC instances include

modification, and when the nouns in LVCs are plural, modification is found in 90% of the cases

observed (1999: 182-3). Although debatable, Kyto (1999), like Matsumoto (1999), finds

evidence that older, more frequent LVCs are more likely to appear without modification of the

noun, indicating that these LVCs may have increased in syntactic fixity over time. Despite this

complex picture, it seems plausible that the main motivating factor for LVCs throughout the

history of English to today is the ease of modification that LVCs offer in comparison to their

lexical verb counterparts.

3.3 Ease of modification

As we have seen, it is so common for some type of modification to exist within LVCs (adjectival

modification, relativization of the noun object) that it is widely noted in the literature on LVCs.

Brinton and Akimoto note,

> "The most salient structural property is the flexibility of verbal modification that [an
> LVC] allows; adjectival modifiers of complex verbs appear to be much easier to use and
> greater in number than adverbial modifiers of simple verbs" (1999: 2).

Yet, despite the salience of this feature, many authors conclude that the main functions of LVCs

is both to displace eventive meaning to an element other than the verb, and convert an atelic

activity into a telic accomplishment, as discussed in the previous section.

This seems to be a trend in the existing research: authors may note the remarkable level

to which modification appears in these structures, but do not seem to consider this a key part of

the function of LVCs in contrast to lexical verbs (Prince, 1972; Live, 1973; Tanabe, 1999),

focusing instead on aspectual functions. However, other authors have proposed the ease and

potential for modification as the primary function of LVCs. When limiting their discussion only

to OE, which did not make use of articles in LVCs, Akimoto and Brinton do admit that

modification appears to be one of the primary motivations for the use of an LVC over a lexical

verb (1999: 50). Matsumoto focuses on ME, and finds that modification and relativization are

important characteristics of LVCs, concluding that "these would seem to be the main motivation

for choosing the CP over the simple verb" (1999:78). Similarly, Kyto (1999), examining early

Modern English, sees the major reason for LVCs' popularity as the flexibility and modifiability

of the structure in comparison to the more rigid lexical verb. Nickel focuses on reanalyzing

complex verbal structures within the framework of generative transformational grammar, and

similarly concludes that "the main motivation...is the fact that the nouns used in complex

structures can be modified in various ways to give expression to infinite shades of

meaning" (1978:77). This leads to the question of whether ease of modification may be the

primary factor for selecting and LVC over a lexical verb counterpart in today's English.

3.4 Corpus-based analysis

Unfortunately, there are no corpus-based, published studies on the level of modification found in

the LVCs of present day English. To address this, the following sections will give an overview

of original work investigating the quantitative and qualitative differences in the modification of

LVCs and lexical verbs in present-day English.

3.4.1 Materials & Methods

The PropBank corpus includes annotations identifying English LVCs and providing semantic

role tags for each of the arguments of the complex predicates, including both syntactic arguments

of the light verb and syntactic arguments of the nominal predicate. In the PropBank corpus,

frequent or more core arguments are assigned numbered arguments while adjuncts or modifier

arguments are assigned an ArgM label. Table 3.4.1-1 gives the possible ArgM labels that can be

used with noun relations, as well as the number of instances of each tag found in a sample of

1,820 LVC annotations. Definitions are drawn from the PropBank annotation guidelines.[9]

PropBank Tag	Role Type	Number of Instances	PropBank Tag, ctd.	Role Type, ctd.	Number of Instances
ArgM-ADJ	Adjectival modification, no semantic specification	377	ArgM-LOC	Location	147
ArgM-CAU	Cause	34	ArgM-MNR	Manner	359
ArgM-COM	Comitative	4	ArgM-NEG	Negative marker	146
ArgM-DIR	Direction, source	12	ArgM-PRD	Secondary Predication	17
ArgM-DIS	Discourse marker	145	ArgM-PRP	Purpose	66
ArgM-EXT	Extent	85	ArgM-REC	Reciprocal	15
ArgM-GOL	Goal	31	ArgM-TMP	Temporal	523

Table 3.4.1-1: Noun relation ArgM types, with number of instances found in LVC annotations.

Unfortunately, these labels do not include tags for a variety of modifier types that may be of

interest for the study of LVCs because they are modifiers that are unique to nouns and can

generally only be expressed periphrastically with verbs. Without a more specific tag, such

arguments are assigned the general ArgM-ADJ tag, which can include any type of modifier or

argument of a noun relation that does not fit one of the other categories, or they are assigned no

tag at all in the case of most determiner types. In order to further study modification within

LVCs, all LVC PropBank instances were further annotated with additional modifier tags by two

PropBank adjudicators, including the present author. The tags were only assigned within the

---

[9] http://verbs.colorado.edu/propbank/EPB-Annotation-Guidelines.pdf

domain of locality of the noun. The additional tags are given in Table 3.4.1-2.

Tag	Role Type	Example
ArgM-CXN	Modifiers included in comparative constructions	He made a **stronger** choice than…
ArgM-REF	Referential determiners	the, this, that, these, those, my, your, etc.
ArgM-QNT	Quantifiers	He made his **third** mistake.
ArgM-ATT	Attributive, Depictive expressions	He did **administrative** work.
ArgM-EVL	Evaluative expressions	He made a **terrible** decision.
ArgM-RLV	Relative clauses	He made the offer **that I was hoping he would make.**

Table 3.4.1-2: Modifier tags developed for the current research, which reflect modifier types that are often used with nouns, but are either incompatible with verbs or can only be used periphrastically with verbs.

A pilot study of modification in the COCA corpus distinguished three different categories of LVCs with potentially differing levels of modification. The three categories differ in the character of the noun denoting an event or state: 1) morphologically related to a lexical verb, or a zero-derived nominal that is identical to a lexical verb (e.g. *make a decision, make an offer*) 2) etymologically related to a lexical verb (e.g. *have a thought, give a speech*), or 3) not clearly related to a verb from a synchronic perspective (e.g. *give an overview, make an effort*). LVCs of category (3) are perhaps the simplest to explain in terms of semantic differences from lexical verbs as they mirror a crosslinguistically common situation: because there is no exact lexical verb alternative in the language that can be used to express the concept, speakers will use an existing eventive noun within an LVC to express the concept instead.   LVCs of category (2), while similar to category (3), are slightly more problematic for an explanation of how LVCs differ from lexical verbs because nearly equivalent lexical verbs exist, which can be used to express the same event. Finally, LVCs of category (1) are especially perplexing: why should an

LVC exist to express a concept that can clearly be expressed (more economically) by a lexical verb alone? For example, "I made a decision to buy the house," vs. "I decided to buy the house."

The pilot study showed that LVCs of category (1) and (2) do tend to have higher levels of modification than LVCs of category (3), motivating the following hypothesis:

Modification Hypothesis: Speakers elect to use an LVC of Category 3 out of necessity — simply because a semantically similar counterpart verb doesn't exist — but speakers use LVCs of Categories (1) and (2), as opposed to their verbal counterparts, because they want to convey additional descriptions of the event, facilitated by nominal expression.

According to this hypothesis, we might expect to see higher levels of modification across category (1) and (2) LVCs, and lower levels of modification across category (3) LVCs. Furthermore, we might expect to see higher levels of modification in LVCs of categories (1) and (2) in comparison to their counterpart lexical verbs. To test the Modification Hypothesis on a larger scale, each of the PropBank LVC annotations was annotated with the LVC category, in addition to the more refined argM labels.

For comparison, the top 20 most frequent LVCs of categories (1) and (2) were extracted. For these LVCs, all of the PropBank annotations of counterpart verbs were also extracted. Finally, for comparison to nouns outside of LVCs, 18,734 annotations of noun relations, not included in LVCs, were also extracted. The result was a corpus of 1,820 detailed LVC annotations with the additional noun modifier tags, a corpus of 10,432 annotated counterpart verbs tokens, and a corpus of 18,734 annotated noun relations.

97

3.4.2 Results & Discussion

The Modification Hypothesis is compatible with the expectation that levels of modification may be higher across LVCs of categories (1) and (2) than category (3), as seen in the pilot study. This finding was not replicated in the full corpus study. The levels of modification were approximately the same across LVC categories, and actually a bit higher across LVCs of category 3: 65% of Category 1 LVC tokens were modified, 64% of Category 2 LVC tokens, and 70% of Category 3 tokens. These results are summarized in Chart 3.4.2-1, which also lists the number (n) of each LVC type found in the corpus.
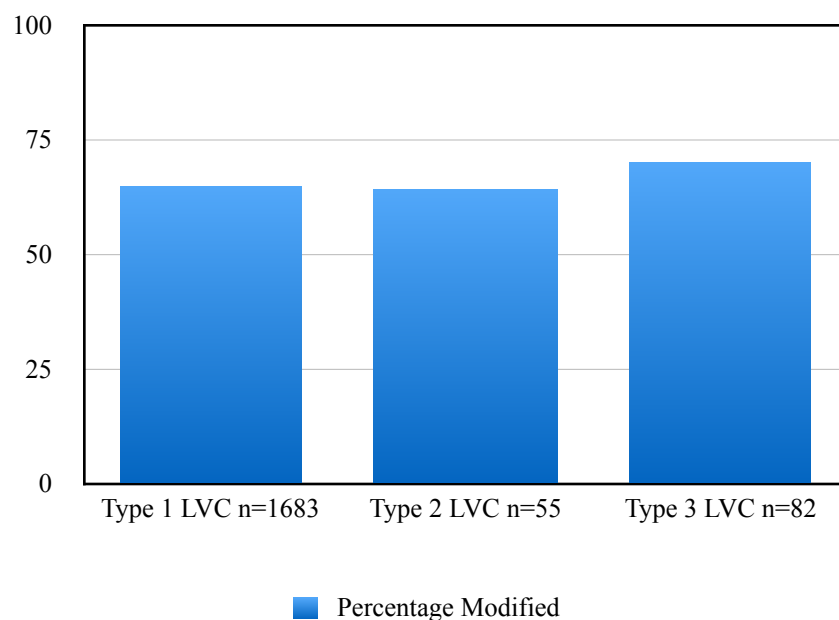
Chart 3.4.2-1: The percentage of LVCs of each type (1-3, based on whether the noun is 1) derivationally related to a verb, or a zero-derived nominal, 2) etymologically related to a verb or 3) not clearly related to a verb) that were modified within the domain of locality of the noun, along with 'n,' the number of each type.

The Modification Hypothesis also predicts that levels of modification will be higher across LVCs of categories (1) and (2) than across their counterpart lexical verbs. This portion of the hypothesis was clearly supported. With just three exceptions out of the 20 cases of LVCs and

counterpart verbs compared, LVCs are more often modified than their verbal counterparts: overall, there are about 1.15 modifiers per LVC relation and 0.60 modifiers per verbal counterpart. Chart 3.4.2-2 shows the average number of modifiers per predicate type, across all predicates of that type. Chart 3.4.2-3 gives the average number of modifiers per LVC predicate and counterpart lexical verb. Additionally, the number (n) of each LVC and counterpart lexical verb examined in the corpus is given.
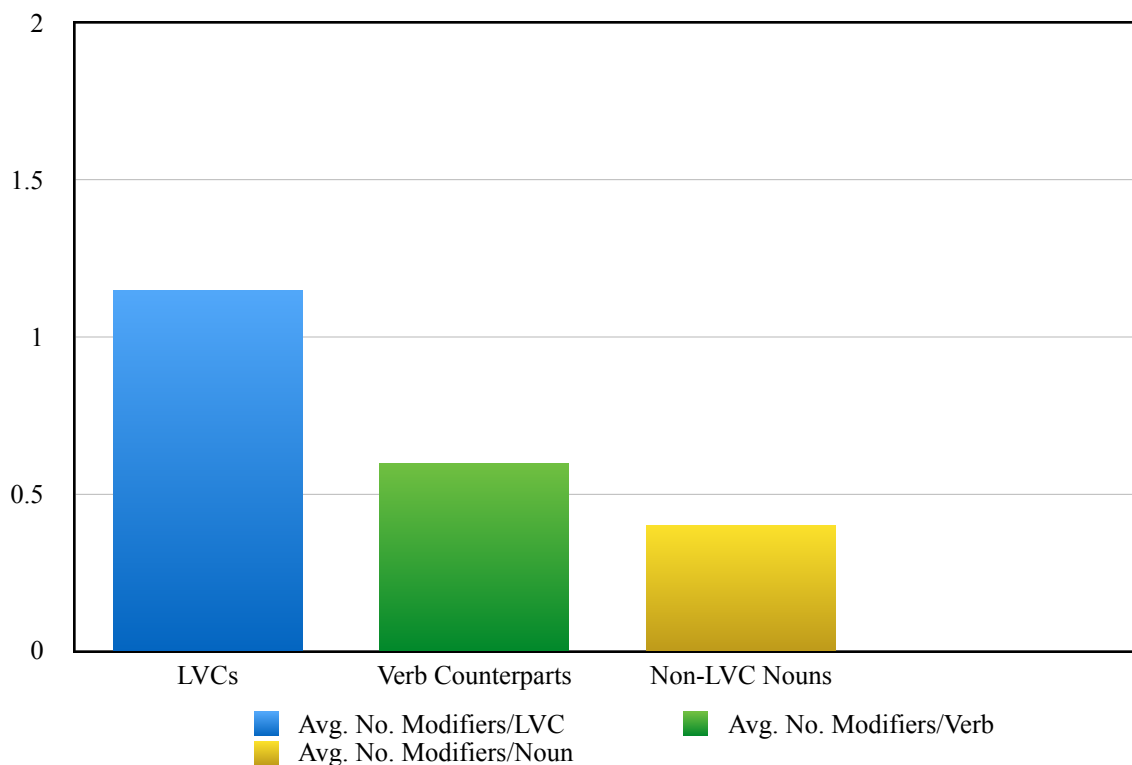
Chart 3.4.2-2: The average number of modifiers, comparing all LVCs with all lexical verb counterparts, and noun predicates outside of LVCs.
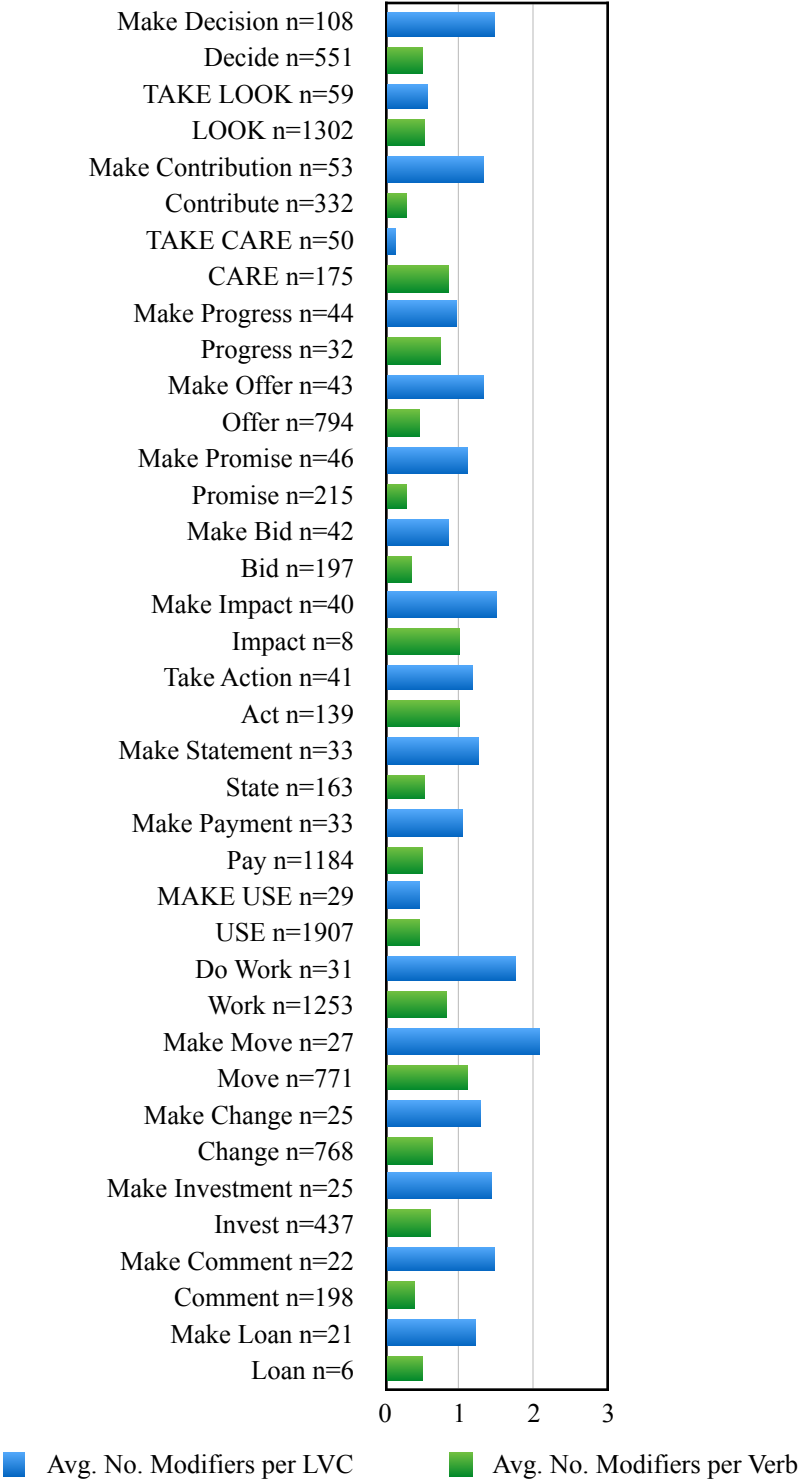
Chart 3.4.2-3: The average number of modifiers, comparing LVCs and lexical verb counterparts. Exceptional cases discussed below are shown in caps.

Furthermore, the levels of modification within LVCs is distinctly higher than that of nouns outside of LVCs. For the 18,734 noun annotations investigated, there were, on average, 0.40 modifiers per noun relation. However, these noun annotations do not include determiners, and 402 of the LVC modifier annotations were referential determiners. If we exclude these modifiers, then the LVCs have, on average, 0.92 modifiers per instance — still notably higher than the level of modification of nouns outside LVCs.

The three exceptions wherein the lexical verb was modified approximately as often, or more often than the LVC were: *look* (average of 0.55 modifiers/token) and *take_look* (average of 0.56 modifiers/token), *care* (average of 0.85 modifiers/token) and *take_care* (average of 0.14 modifiers/token), *use* (average of 0.45 modifiers/token) and *make_use* (average of 0.45 modifiers/token). These exceptions have reasonable explanations when the data is explored. Firstly, *take_care* and *make_use* are somewhat exceptional LVCs because they generally do not take any article, and for this reason they have likely developed a stronger cohesion over time, discouraging intervening modifiers, as suggested in the previous sections on diachronic change.

As exemplified in the title of this dissertation, *take_look* (as well as *take_care*) is often used imperatively within commands or requests, thus speakers generally do not need to convey rich event descriptions, since the event is hypothetical or future. This brings up an important point about the function of LVCs generally: while this research attempts to find an overarching reason why LVCs exist alongside counterpart lexical verbs, there are certainly unique motivations for speakers to use individual LVCs in a particular context, and as a particular LVC becomes more commonly invoked in that context, then it can become the less marked, more formulaic expression in that context (Wierzbicka's (1982) work nicely exemplifies some of these

cases). For example, *take a walk* is largely used for pleasurable, aimless walking, while *walk*

may be invoked for goal-oriented walking. This research is not meant to discount these

explanations, but because they are somewhat idiosyncratic to the semantics of the LVC, the

counterpart verb in question, and their frequency of use in particular contexts, such explanations

are impractically time-consuming to detect for all LVCs. Additionally, as Wierzbicka's work

also demonstrates, it's likely that the results of such an analysis of differences in meaning are

somewhat subjective, and it seems unlikely that they could be reproduced. This analysis instead

turns to empirical facts about the usages of LVCs and counterpart lexical verbs to demonstrate a

general function of LVCs as opposed to counterpart verbs.

Modification is not simply more frequent in the case of LVCs than lexical verbs, it is also

qualitatively different. In English, nouns can be modified not only by adjectives, but also by

other nouns, in the case of noun-noun compounds, and by relative clauses. Additionally, nouns

are compatible with various types of determiners, quantifiers and articles (such as possessive

pronouns and cardinal numbers) that can add shades of meaning, including those of an aspectual

nature, or can contribute to reference tracking. Although there is some overlap in what can be

expressed by nominal modification and what can be expressed by verbal modification (e.g. the

manner of the event, *deliberate* vs. *deliberately*), much of the modification within LVCs has no

counterpart in verbal modification, and could only be expressed periphrastically. This can be

observed in the following usage examples of LVCs with evaluative or attributive modification,

followed by an invented example that attempts a reformulation of the LVC using the

corresponding verb:

83. We had -- we had a [**really good**]-*evaluative* laugh and then things happened.

84. ?We laughed **really well** and then things happened.

Example (84) shows a typical alternation in meaning: the transformation of an evaluative

nominal modifier into an adverbial modifier changes the meaning of the modifier into what is

most likely to be interpreted as a somewhat awkward and ambiguous manner modifier. At best,

we could rephrase this as *We laughed really hard and then things happened,* but even this

changes the meaning of the phrase. Other nominal modifiers, seemingly primarily those that are

evaluative, simply do not have an adverbial equivalent that is both grammatical and semantically

very similar:

85. He did extremely well raising money on the Internet and gave [**one of the more
    interesting**]-*evaluative* speeches.

86. ?He did extremely well raising money on the Internet and spoke **more interestingly than
    others.**[10]

87. President Bush gave a [**very big**]-*evaluative* speech in September of that year -- transferring
    those detainees to Guantanamo.

88. *President Bush spoke **very bigly** in September of that year -- transferring those detainees to
    Guantanamo.

The qualitative differences seen in the types of modifiers of LVCs and the types of modifiers of

the corresponding verbs provides further evidence for the reason that LVCs of type (1), which

have morphologically equivalent lexical verbs, exist alongside these verbs: LVCs provide a

---

[10] While this sentence is grammatical, it doesn't seem to carry the same meaning as the LVC alternative. Here, "interestingly" seems to convey the manner of speech (perhaps an interesting accent), rather than serving as a comment on the content of the speech.

resource of the language, or coding means, exploited by speakers to express an event as a noun, and therefore also express descriptions of the event using nominal modification, which is more flexible than verbal modification and allows for the grammatical and economical expression of descriptions.

3.4.3 Conclusions

When comparing views on the function of LVCs, a picture emerges wherein authors researching historical English primarily point to the modificational potential of LVCs as their function, while authors researching present-day English (with the exception of Nickel, 1978), tend to point to the aspectual function of LVCs, and specifically to usage of the singular, indefinite article, which is taken to be evidence of telicity. These tendencies are consistent with the development of article usage in the English language. In OE and ME, when the usage of articles was not fully developed or consistent, there does not seem to be a clear aspectual function of LVCs. However, as article use has developed to the present day, the singular, indefinite article within LVCs now seems to lend a bounded, countable nature to the event, making LVCs seem somewhat telic or perfective in nature. In reality, these two views on the function of LVCs, emphasizing either the importance of modification or aspectual qualities, are not at odds. Articles lending telicity are merely one of the many elements compatible with nouns that can be used within LVCs to express subtle meanings. Given that LVCs have been used for their modificational potential throughout the history of English, it is not surprising that speakers would use LVCs in combination with articles and pluralization in order to convey certain aspectual nuances once these resources had become fully developed in the English language. However, a full examination of these views

demonstrates that LVCs are used with many types of modificational elements, of which those

that lend telicity are only a subset.

This view is compatible with Horn's (1985) Division of Pragmatic Labor, a principle that

was developed to account for similar, competing linguistic phenomena observed by Horn.

Specifically, Horn noticed other forms existing in English that seemed to violate Gricean

Maxims of Quantity, and discusses why these competing, arguably less efficient, forms were not

blocked or eliminated from the lexicon. For example, Horn notes cases where the existence of a

simple abstract nominal underlying a *-ous* adjective blocks or prevents the formation of an *-ity*

nominalization based on the adjective, for example: *fury, furious, *furiosity.* However, this does

not occur in all cases, as exemplified in: *glory, glorious, gloriousness.* Horn therefore

hypothesizes that such forms will not be blocked if they pick up some restricted meaning, and

will be blocked where there is no meaning "left over" for the more productive form to pick up

(1985: 26). Similarly, Horn points out that periphrastic causative constructions may exist

alongside a lexical alternative (much like LVCs); for example:

89. Black Bart killed the sheriff.
90. Black Bart caused the sheriff to die.

In this case, Horn hypothesizes that the use of the morphologically more complex, and therefore

marked, causative form implicates that the unmarked situation did not occur (perhaps accidental

"killing" occurred instead), while the use of the simple lexical version implicates that the action

was brought about in an unmarked way. From these observations, Horn develops his notion of

the Division of Pragmatic Labor:

> "Use of a marked (relatively complex and/or prolix) expression when a
> corresponding unmarked (simpler, less effortful) alternate expression is available
> tends to be interpreted as conveying a marked message (one which the unmarked
> alternative would not or could not have conveyed)" (1985: 22).

In the domain of LVCs, this notion is arguably exemplified in two somewhat distinct manners.

Firstly, speakers use the longer, marked LVC form to convey messages with detailed event

descriptions that cannot be conveyed by the verbal, unmarked form. In these cases, the fact that

speakers are motivated to express such descriptions could be thought to stem from the marked

nature of the event (generic, unmarked events likely do not need any further description).

Secondly, it's important to note that over time, the LVC may become the conventionalized

expression for a marked event. For example, *make_offer* is associated with certain types of

offering events (arguably more formal), and may not be appropriate in its application to other

offering events. This is exemplified by the fact that it seems a bit strange to use the LVC in

everyday, casual offering events:

91. I offered ice cream to the child.

92. ?I made an offer of ice cream to the child.

93. I made him a generous job offer.

94. ?I made him a generous ice cream offer.

Thus, LVCs can take on a more specialized meaning over time. However, pinpointing the

meaning that an LVC picks up in comparison to its counterpart lexical verb, as mentioned

previously, is perhaps an intractable process for all LVCs. Nonetheless, the use of the marked

form to denote marked situations is the reason that the LVC form is retained in English alongside

semantically similar lexical verbs.

Chapter 4
LVC Productivity & Acceptability (Research Question 3)

Closely related to the problem of LVC detection is the issue of LVC (semi-)productivity. LVCs are productive enough that novel constructions can enter the language, making a words-with-spaces approach to LVCs intractable. However, this productivity is not complete — there are constraints on which eventive or stative nouns can acceptably combine with a given light verb. Notably, LVCs tend to occur in semantically similar "families." In other words, semantically similar complements tend to co-occur with the same light verb (and perhaps not with another light verb: *give a cry, moan, howl,* but **take a cry, moan, howl* (Stevenson, Fazly & North, 2004)). Even combinations for a particular light verb seem to be somewhat idiosyncratic in their acceptability. For example, a variety of nouns denoting communication events combine with *make:*

95. make a speech, declaration, proclamation, announcement

96. make a recommendation, suggestion

but other semantically similar nouns may not be acceptable combinations with *make:*

97. ?make a yell

98. ?make advice

One might argue that the questionable extensions suggested here do not make use of the right kind or level of semantic similarity to attested examples, and indeed some research has tried to pinpoint the exact semantic constraints that make a given combination acceptable (e.g. Wierzbicka, 1982; discussed in Chapter 1). This chapter focuses on research relating to LVC

productivity and acceptability, for if we could better understand why certain combinations are acceptable and others are not, and how speakers extend the LVC construction to novel combinations, this would greatly facilitate detection of low-frequency and previously unattested LVCs in NLP.

This dissertation research focuses on the roles of frequency and semantic similarity in the productivity of LVCs. Specifically, the hypothesis that novel LVCs will be acceptable if they are semantically similar to an existing, highly frequent exemplar LVC. This draws heavily on the work of Bybee (2010), who emphasizes the importance of frequency and analogy in extending constructions generally, and Bybee and Eddington (2006), who examine this hypothesis as it applies to Spanish *becoming* constructions, which are similarly idiosyncratic in their productivity. This approach to productivity assumes the importance of frequency in language; thus, the chapter begins with an overview of Emergent Grammar, which assumes that grammatical structure arises from repetition. The chapter then turns to a more in-depth examination of frequency effects in grammar. We'll then examine linguistic and computational approaches to handling the (semi-)productivity of constructions, and the chapter culminates in a description of human judgments of acceptability that test this hypothesis.

4.1 Background: Building the Hypothesis

This research is couched in a usage-based, emergent view of grammar. In this view of grammar, the same domain-general processes that operate in other areas of human cognition also operate in the domain of language. These cognitive processes operate in each instance of language use, and the repetitive use of these processes have an impact on the cognitive representation of language

and overt language use (Bybee, 2010). As a result, linguistic structure "emerges" from use (Hopper, 1987). As a mental structure that is in constant use, and is filtered by processing activities that change linguistic representations, there is much variation and gradation in language, as opposed to discrete categories. It is for this reason that we see a variety of continua in language (such as the continua of compositionality and idiomaticity) instead of binary distinctions.

One prerequisite for this view is a rich memory for each token of linguistic experience, or each 'exemplar.' These representations contain information concerning phonetic details, lexical items and constructions used, meaning, inferences made from this meaning and context, as well as the properties of social, physical and linguistic context. On this point, usage-based theories are quite different from the structural and generative views of linguistics. In those views, speakers do not need to store information on each linguistic exemplar (and indeed, many thought that this type of memory was not possible for humans) (Jakobson, 1990). Instead, speakers abstract away from specific instances and find the right information to characterize a pattern as a rule. Only idiosyncratic information needs to be stored with lexical items, otherwise speakers can rely on knowledge of the abstract grammar rule for interpretation. Advances in our understanding of the human brain and capacity for memory have shown that such rich memory is plausible. Gurevich et al. (2010) challenge previous studies showing that speakers only remember "the gist" of utterances (Sachs, 1967; Bransford & Franks, 1971; cited in Gurevich et al.), and demonstrate that speakers can remember verbatim phrases from stories 72% of the time, and when retelling stories, speakers use from 9-22% matches to clauses in the original story.

Under this "exemplar view," frequency of use plays an important role. Each token of linguistic experience strengthens the representation of a particular word, string, or construction (Bybee, 2006). Thus, high-frequency exemplars will be stronger than low-frequency exemplars, in the sense that high-frequency exemplars are easier to access and show increased stability, or are less susceptible to change. Finally, more frequent exemplars tend to serve as the basis of new forms, extended by analogy to existing forms.

One important question for the usage-based view is how new forms emerge, if grammar is based solely upon experience with particular exemplars. New forms are created based on analogy to previously experienced utterances. In this process, novel utterances are formed from previously experienced utterances in a process where the novel usage is compared to, aligned with, and categorized as a member of a cluster of similar utterances. The novel forms can then take on unique, independent meanings as they are used with some frequency in, and become associated with, a certain context (Bybee, 2010). Bybee cites a frequently studied example, the WXDY? construction (Fillmore and Kay, 1999; Johnson 1997), shown here in a famous joke:

Diner: *Waiter,* ***what's this fly doing in my soup?***

Waiter: *Why, madam, I believe that's the backstroke.*

(Fillmore and Kay, 1994)

The bold-faced portion of this exchange is an example of the WXDY? construction. The joke itself plays upon the ambiguity between the literal interpretation of this phrase and the constructional meaning. The waiter responds as if the expression is a literal request for information, whereas, of course, listeners who find the joke to be funny understand that the expression indicates surprise and disapproval over the fly's very presence in the soup, and is not

a request for information about what the fly is up to. The important question for this research is how this additional meaning comes to be associated with such a general Wh-question involving only the verb *doing* and a locative phrase. Bybee (2010) asserts that these implications arise from language use in context. The implications of surprise and disapproval must have come from multiple instances of use in contexts that invoke this negative nuance. As mentioned previously, the validity of this proposal relies upon speakers recording in memory inferences drawn from a variety of linguistic levels, including social context. As a particular construction is used with greater frequency in a particular context, such inferences can become part of the meaning of the construction (Bybee, 2010).

Thus, new constructions arise out of specific exemplars of old constructions. Constructions themselves have some fixed slots and some open slots, allowing for further variation and extension. For example, the WXDY? construction has been extended to a variety of conventionalized phrases akin to *What's a nice girl like you doing in a place like this?,* perhaps firstly or most notably used in the 1953 film *The Wild One*. The open slots in the WXDY? construction seem to be fairly permissive of a variety of noun phrases (NPs) and locative expressions, but other constructions have more nuanced constraints on what type of items are permissible in the open slots. Furthermore, determining the appropriate type of semantic element that can sit in an open position can be difficult; for example, although the WXDY? construction is often specified as involving a locative preposition (PP), arguably other PPs are permissible and entail the same meaning (e.g. *What are you doing **with that knife**?!*). Therefore, the productivity of a certain construction is determined by the range of elements that can acceptably fill an open slot.

It is difficult to define LVCs in terms of fixed and open slots in the same way that we can loosely define the WXDY? construction. In order to capture the appropriate level of generalization and constraints, we cannot assume that the "construction" is any *light verb + noun* combination. Instead, usage-based approaches and the exemplar model would suggest that the "construction" extended in this case is a particular light verb in combination with certain members of a semantic class of nouns. This interpretation is supported by the fact that we do see semantically similar "families" of LVCs, where the verb remains the same and similar nouns felicitously combine with that verb. Thus, the open slot, with tricky constraints upon what can acceptably fit there, is that of the noun.

Under this view, productivity is determined on the basis of acts of categorization, wherein novel constructions are compared to familiar constructions and potentially categorized as an "acceptable" member of a family of existing constructions. The primary factors of this process, frequency and categorization, will be discussed in greater detail in the sections to follow.

4.1.1 The Role of Frequency

Some of the effect of frequency on grammar has already been described in the previous section. To summarize, frequency is thought to 1) "strengthen" a representation by making frequent forms easier to access (and thereby likely making them more frequent still), 2) increase the stability of a form by making it more resistant to change, or more entrenched and ossified (a notable exception to this is the reduction of words due to very frequent, formulaic use: e.g. *I dunno, I'm gonna*), 3) make a form more likely to serve as the basis for extension of a form by

analogy, 4) enable meaning change as a form is used more frequently in a particular context, and becomes associated with that context.

A plausible corollary to these facts is that there would be a clear relationship between frequency and acceptability: frequent forms are more acceptable, while infrequent forms may be less acceptable. Indeed, it is tempting to think that perhaps speakers recognize, for example, a verb's nonoccurrence in an argument structure, and immediately rule it out for acceptable use in that construction (Goldberg, 2011). In other words, only formulations heard with some regularity are acceptable. However, there are many examples in language providing evidence against this view. For example, many verbs that a speaker has likely heard frequently in the intransitive form can acceptably occur in certain "coercive" constructions: *He sneezed the milk out of his nose* (Goldberg, 2011). Thus, we can begin to see that there is not a straightforward relationship between acceptability and frequency.

Bermel & Knittl (2012a) have examined the relationship between corpus frequency and acceptability in some detail. The authors note two important forces or assumptions in the Emergent Grammar view: entrenchment and statistical preemption. Entrenchment is the process whereby forms and constructions are fixed through repeated exposure and usage, and thereby become the basis of linguistic structure. Statistical preemption is the process whereby forms achieve predominance in usage over competing entities, and render potential competitors unacceptable to speakers. Divjak (2008) brings up a potential argument against the Emergent Grammar view: this view cannot explain why some constructions continue to be used at low levels of frequency relative to more frequent alternatives (i.e. why are there competing "forms in

variation"). With these forces and issues in mind, Bermel & Knittl examine three hypotheses

concerning the relationship between corpus frequency and acceptability:

> **Strong hypothesis**: data from the corpus on the relative proportions of forms in
> variation should reflect in rough measure the acceptability of these forms to native
> speakers.
>
> **Weak hypothesis**: there is a correlation between corpus data and acceptability,
> but it is not proportional or symmetric, i.e. we cannot count on the proportions to
> correspond precisely to value judgments, nor is it always possible to abstract
> predictions about acceptability from corpus data.
>
> **Null hypothesis**: no relation between proportionality in a corpus and acceptability
> of features to native speakers.
> (Bermel & Knittl, 2012a: 245)

The authors test these hypotheses in the realm of Czech variants of nominal declensions,

focusing on cases where two competing, variant forms exist to express the genitive singular and

the locative singular. They select test words that alternate these endings in varying proportions.

To test acceptability of each variant, the authors ask for a scalar judgment, with anchor points 1

and 7, and no descriptors for points 2-6. Two surveys are distributed, with 143 and 136

responses each.

The authors firstly find that higher frequency of one variant (occurring over 50% of the

time in comparison to the other variant) in the corpus does entail a high acceptability rating.

However, a high acceptability rating does not entail a higher frequency in the corpus, and a lower

frequency in the corpus (occuring less than 50% of the time in comparison to the other variant)

does not entail a low acceptability rating. Finally, a low acceptability rating does entail low-

frequency in the corpus. From these findings, the authors conclude that the relationship between

frequency and acceptability is not straightforward, and can only allow us to make some general

assumptions:

> "the relative proportion of a form in the corpus allows us to draw general
> conclusions about its acceptability in relation to other forms, but it does not allow
> us to pin actual values to the acceptability of that form" (2012a: 265).

If there is a situation in which form A has a greater proportion of representation in a corpus than

competing form B, then we can only conclude that B cannot be significantly MORE acceptable

than A. The authors also establish several frequency bands of interest. They find that if a form is

used in 50-100% of the usages of either variant, then they can assume that the form will have

higher acceptability. If the form occurs in 1-49% of the usages, then no prediction can be made.

If the form occurs in 0% of the usages, then low acceptability can be predicted. Bermel and

Knittl conclude that these findings are consistent with the Weak Hypothesis: overall the effect of

frequency is strong, but in individual instances its applicability is limited.

Addressing Divjak's (2008) point, the authors also note that statistical preemption does

not happen reliably: assigning a high acceptability rating to one variant does not imply that a

competing variant gets a low rating. In a second study (Bermel & Knittl, 2012b), the authors

further examine the role of syntactic context in acceptability and the question of why competing

variants are maintained, instead of statistically preempted, in Czech. They find that minority

variants are more strongly associated with a particular syntactic context, wherein the variant is

highly acceptable. For this study, Bermel and Knittl therefore propose a picture wherein

speakers firstly have recourse to their sense of frequency and entrenchment of individual forms,

and secondly rely upon coarser-grained frequency of syntactic structures, therefore the

association of certain nominal declensions with easily abstractable syntactic structures bolsters

their acceptability ratings. Similarly, Langacker (2000:16) points out that while frequency seems

to be the main determinant as to which variant will be used, contextual priming can override the

effects of frequency. Thus, the specialized relationship between a minority form and a particular

syntactic context can prevent that minority form from being preempted, and contributes to

maintaining a linguistic system with variation. This is precisely what is seen in some cases of

LVCs, where, for example, speakers use *take a walk* in the context of pleasurable, aimless

walking where the syntax and semantics do not necessarily include a goal argument.

The forces of entrenchment and preemption have also been explored from the perspective

of psycholinguistics. For example, Boyd, Ackerman and Kutas (2012) attempt to determine

whether grammatical constraints that prevent **the joke giggled me* are acquired through 1) innate

language biases, 2) constraints learned from input via cues from semantics, or 3) statistical

constraint learning, including both entrenchment and preemption. Unfortunately, as the authors

point out, teasing apart preemption and entrenchment is very difficult because they are highly

correlated. To circumvent this difficulty, the researchers use nonsense verbs, presenting subjects

with images of actions and either an intransitive sentence (*Look! The apple is yadding!*), or both

an intransitive and a periphrastic causative sentence (adding *The squirrel really made the apple*

yad!). The periphrastic causative should provide a preempting structure, blocking the use of the

nonsense verb in a transitive construction. The subjects are then asked to produce a sentence

using the same nonsense verb describing the images, and then rate sentences using the verb in a

transitive and periphrastic causative sentence on a five-point grammaticality scale. The authors

find that the "mixed" group members, who were presented with both the intransitive and

periphrastic causative sentences, were less likely to produce a transitive description of images, and rated the transitive sentence as significantly lower than the other group. This demonstrates that preemption operates independently from entrenchment: although both groups were provided with evidence about the verb's use in the intransitive (thereby supporting learning through entrenchment), the use of the periphrastic causative preempts the possibility of using the transitive for the mixed group.

Theakston (2004) also concludes that entrenchment operates independently from preemption in a task where both children and adults are asked to provide grammaticality ratings for sentences with argument structure errors, involving both high and low-frequency verbs. She finds that frequency plays a strong role: for both children and adults, lower-frequency verbs were found to be more acceptable than high-frequency verbs in the erroneous argument structures. Theakston notes that the adults would certainly be expected to have preempting structures for both high and low-frequency verbs, yet this did not remove the frequency effect observed in participants' judgments of grammaticality. Thus, we can see that preemption is not an all-or-nothing process, speakers may remain less confident of the range of constructions in which lower-frequency items are used.

Frequency also plays another notable role in grammar mentioned in much linguistic research: frequency can contribute to semantic bleaching. In Hopper & Traugott's (1993) discussion of grammaticalization, it is noted that as forms become more frequent and entrenched in their use, they undergo semantic bleaching or loss of the original (specificity of) meaning. Within idiomatic and semi-idiomatic phrases, entrenchment may result in semantic bleaching that weakens the strength of the relationship between the phrase and the semantics of its

118

component parts. In other words, it's possible that the idiomatic meaning is directly accessed for frequent, entrenched expressions, and the phrase is never decomposed into the semantics of its component parts.

4.1.2 Productivity, Acceptability and Categorization Models

The acceptability of a particular construction may be based on whether or not speakers are able to categorize that construction within a larger class of previously experienced constructions. In the case of LVCs, speakers and hearers will categorize a candidate combination of *light verb + noun* as either an acceptable member of an existing category of constructions, or an unacceptable construction that does not fit with existing categories. Indeed, research on semi-productive constructions often focuses on how speakers categorize certain constructions as acceptable, while others are categorized as unacceptable and are not used (Bybee, 2006; Bybee & Eddington, 2006; Suttle & Goldberg, 2011). Current proposals suggests that frequency of a certain construction and semantic similarity to frequent constructions both play central roles in structuring speaker's categories of constructions (Bybee 2006; Bybee & Eddington, 2006; Suttle & Goldberg, 2011). These proposals draw on more general models of categorization, which often fall into two opposing camps: the prototype model and the exemplar model.

Each of these models are distinct from the classical, Aristotelian view of categorization, wherein entities can be classified according to common, critical features resulting in categories with clear boundaries (Lakoff, 1987: 8). In the prototype model, categorization involves a process of abstraction, where central tendencies of exemplars' features are used in a summary representation, or prototype of the category; novel instances are then categorized based on

similarity to the prototype representation (e.g. Posner & Keele, 1970; Rosch & Mervis, 1975). In the somewhat opposing view, the exemplar model, the process of categorization relies on specific knowledge of representations of earlier instances or exemplars; novel instances are classified based on their similarity to earlier instances (e.g. Brooks, 1978; Hintzman, 1986; Medin & Schaffer, 1978). An important distinction in these models is that the prototype view requires the formation of an abstract representation through analysis of the features of exemplars, whereas in an extreme exemplar view, each instance of a category is stored as a relatively unanalyzed whole.

More recent and compelling research into categorization proposes a combination of these models: categorization relies upon the storage and use of both abstract, general information involved in a prototype representation, as well as specific information about individual exemplars (Malt, 1989; Ross, Perkins & Tenpenny, 1990). Furthermore, Barsalou (1985) proposed that categories involve graded structure: a continuum from most representative category members to most atypical category members, and finally to nonmembers least similar to its category members. Barsalou's research into the determinants of graded structure demonstrated that categories don't seem to have invariant structure, but instead can be dynamically constructed based on the context and type of categorization involved, with special attention to the distinction between common taxonomical classification (e.g. what animals are mammals) and goal-oriented classification (e.g. what items should one take on a camping trip). In the online, dynamic construction of categories, various determinants of graded structure can be emphasized or attenuated depending on the context and type of categorization. Specifically, Barsalou found that taxonomical classification relies more heavily on an understanding of a

category's central tendencies, but goal-oriented categories rely on an exemplar's similarity to an ideal category member. Barsalou also notes the importance of frequency of instantiation, which Barsalou defines as "someone's subjective estimate of how often they have experienced an entity as a member of a particular category" (1985:631). This suggests that perhaps the frequency with which a given verb and noun occur across all LVCs may have an impact on the acceptability of a particular LVC combination (Stevenson, Fazly & North, 2004 and Fazly, North & Stevenson, 2005 draw upon this intuition in their computational measures of productivity and acceptability, discussed in Section 4.2).

The possibility of employing different strategies leads to the question of when one categorization strategy is used over another. Malt (1989) used a priming paradigm to provide evidence that subjects invoke stored exemplars in certain categorization tasks, but seem to invoke a prototype in others. Although inconclusive, subject reports indicated that subjects used a prototype strategy to categorize typical stimuli and used an exemplar strategy to categorize atypical stimuli.

Research into semi-productive linguistic constructions suggests that a combination of an exemplar model and a prototype model are used to construct the categories of constructions in grammar (Bybee, 2006; Bybee & Eddington, 2006). Specifically, Bybee (2006) suggests that speakers categorize each instance of a construction that they have heard in their experience. When speakers then hear a specific construction, they are mapped onto identical existing representations if the construction has been heard before, and this strengthens the representation of that construction. It is this representation that is an exemplar in this model, stored in the speaker's mind as an unanalyzed whole (Bybee & Eddington, 2006: 1; Malt, 1989). If the

construction has not been heard before, speakers compare the novel construction to existing

representations, and judge the novel construction as similar or dissimilar to existing exemplars

(Bybee & Eddington, 2006: 1). Similar exemplars then cluster together to form categories that

have a prototype structure; the prototype itself is the most frequent exemplar of that cluster.

Therefore, in this view, speakers categorize a new construction as acceptable or unacceptable

based on 1) whether they have heard the construction before, 2) whether it is semantically similar

to a construction they have heard before. Creative or productive usages of a construction can be

formed by analogy to existing exemplars (Israel, 1996; Bybee & Eddington, 2006).

Proponents of this approach to understanding constructions argue that a strict prototype

model is not adequate because a prototype model would require speakers to predict which

features of an exemplar will be chosen in the future as the basis for extending the construction

(Bybee & Eddington, 2006: 6). This requirement stems from the nature of the prototype model,

wherein categories are created by extracting features from instances, but full exemplars are not

stored in memory (Rosch & Mervis, 1975; Malt, 1989). With respect to LVCs, the exemplar and

prototype models would differ concerning what type of information is stored about action

nominal complements that can combine with a particular light verb:

Exemplar

take + walk, stroll, run

Prototype

take + nominals with features X, Y, Z

In the exemplar model approach, the LVC could be extended based on any feature of stored

exemplars (because all features are stored), whereas in the prototype model approach, speakers

would only extend an LVC based on certain salient features of the prototype.

A second argument for an exemplar model of constructions is the importance of frequency in acceptability judgments of constructions. Research has shown quite conclusively that constructions found with a high-frequency in a corpus, as well as constructions that are very semantically similar to highly frequent constructions, are generally judged as acceptable. This is thought to show that frequency of occurrence is somehow encoded in the mind, and acceptability is assessed on the basis of similarity to exemplars (Bybee & Eddington 2006; Suttle & Goldberg, 2011). From the perspective of a prototype model, exemplars are not stored, so speakers clearly could not be attuned to the frequency of exemplars. However, arguably frequent features could be those that make up a prototype. It is important to note that this research does not rule out the validity of a prototype model (as exemplars cluster into categories around a prototype); however, this research hinges to some extent on speakers' storage of whole instances, rather than some certain features of instances.

Suttle and Goldberg (2011) also find that frequency and semantic similarity are important aspects of judging acceptable and unacceptable LVCs; however, they also point out the importance of several other factors. First, they argue that novel extensions of a construction are more likely to be judged acceptable if the pattern of that construction is relatively variable. Essentially, speakers will consider a particular construction as extendable if they have often witnessed the pattern being extended in various ways. Secondly, they argue for the importance of *coverage*: the degree to which attested instances fill the semantic/phonological space that includes the potentially novel usage. Coverage is not directly related to type frequency or variability; rather it is determined by the relationships between attested instances and the target usage in question. As an example illustrating the concept of coverage, Suttle and Goldberg refer

to an experiment where speakers are more confident of the conclusion in (A) than the conclusion

in (B):

(A)   assumption 1: Lions have property X.
       assumption 2: Giraffes have property X.
       conclusion: Rabbits have property X.

(B)   assumption 1: Lions have property X.
       assumption 2: Tigers have property X.
       conclusion: Rabbits have property X.

The authors propose that coverage is the main explanation for the preference for (A): (B) tells us

something about the category of large felines, which does not include rabbits, and (A) leads to a

generalization that may hold for all mammals. According to Suttle and Goldberg, the importance

of coverage can explain the opposing interaction between variability (where variability is greater,

a construction is more extendable) and similarity (a construction is extendable to the extent that

is similar to an attested instance).

In summary, research on how speakers construct and extend categories of acceptable

constructions suggests the importance of the following factors:

1) Frequency of a particular construction
2) Semantic similarity to an attested, frequent construction
3) Variability of the construction
4) Coverage: novel usages are acceptable to the extent that semantic/phonological space
   is well covered by the smallest category encompassing both the novel usage and
   attested constructions.

However, the importance of these factors has been demonstrated primarily in research on other

types of constructions besides LVCs. Thus, it is important to examine whether or not these

factors are truly important for judging acceptability of all types of constructions, or merely the

types of constructions already investigated.

4.1.3 A Case Study: Spanish *Becoming* Constructions

In Bybee and Eddington (2006), the productivity of Spanish *becoming* constructions is analyzed. Spanish *becoming* constructions are in many ways very similar to English LVCs, given that several verbs roughly meaning *become* (e.g. *ponerse, volverse*) combine idiosyncratically with certain adjectives; for example, *volverse loco* ('go/become crazy') is acceptable, but *?ponerse loco* is marginal. Alternatively, *ponerse pesado* ('become annoying') is acceptable, but *? volverse pesado* is marginal. Furthermore, as discussed in Chapter 1, copular verbs like *become* do constitute a subtype of support verbs. In the case of Spanish *becoming* constructions, Bybee and Eddington assume an exemplar model, and further assume that high-frequency exemplars would be judged more acceptable than the low-frequency exemplars that may not be robustly represented in linguistic memory. The authors also note that the less frequent exemplars were often cases where the adjectives were synonymous with adjectives that occurred in conventionalized expressions, leading them to imagine a model wherein the conventionalized adjectives constituted the central members of categories that fanned out from these centers. This model seems highly plausible for English LVCs as well.

Bybee and Eddington test the validity of this model by firstly grouping together adjectives occurring with four roughly synonymous Spanish verbs in a corpus of 1.1 million words of spoken Spanish and a second corpus of novels consisting of 990,000 words. The semantically similar groups were checked by asking a single native speaker of Spanish to perform a sorting task with the adjectives, and these were further tested by having a second set of seventy-seven speakers provide ratings for the semantic similarity of certain pairs of adjectives.

Thus, the adjectives were grouped by similarity of meaning. For example, the adjective *solo*

('alone') is grouped with the adjectives *soltera* ('single, unmarried') and *aislado* ('isolated').

The authors could then examine the adjectives that were semantically similar to those that

occurred multiple times with certain verbs. This allowed for the detection of semantically

similar families, or clusters, where, for example, *quedarse solo* occurs 28 times in the corpora

and *quedarse aislado* occurs just twice. The authors provide a detailed analysis of each of these

clusters.

   After this clustering experiment, Bybee and Eddington perform an acceptability

experiment to test the role of token frequency along with semantic similarity in determining

category membership. Specifically, the authors hypothesize that stronger, more frequent

exemplars serve as the basis for the production of novel expressions that are semantically similar

to the higher-frequency exemplars. A corollary is that speakers will judge higher-frequency

combinations to be more acceptable than lower-frequency combinations that do not

(semantically) resemble the frequently used combinations. To test this hypothesis, the authors

develop a questionnaire wherein subjects are asked to rate combinations in sentences on a Likert-

type scale extending from 'Perfectly fine' on one end of the spectrum to 'Odd' on the other end.

There were three types of test items: 1) sentences with high-frequency adjectives that make up

the core of exemplar clusters, 2) sentences with low-frequency adjectives that are semantically

related to one of the core clusters of each verb, and 3) low-frequency adjectives that are

unrelated to other adjectives in the corpora. The questionnaire was presented to 48 native

Spanish speakers.

The second experiment established that there is a statistically significant correlation between token frequency and rating: high-frequency items were generally rated to be 'perfectly fine.' Furthermore, low-frequency items that were semantically similar to high-frequency clusters were rated as 'perfectly fine' significantly more often than the low-frequency items that were not semantically similar to high-frequency exemplars. The authors take this as evidence that these Spanish support verb constructions are extended by analogy to a frequent exemplar. Specifically, the authors envision a process in which speakers hear an unfamiliar construction, and its features are compared to those of an existing cluster of exemplars. Within that cluster, the representation of the most frequent exemplar is strengthened through repetition, and therefore becomes the central member of the cluster. As the central member, the most frequent exemplar becomes the primary basis for comparison when evaluating whether or not an additional, unfamiliar construction should be added to that family of constructions, and deemed acceptable. The validity of this vision in the realm of English LVCs will be examined thoroughly in Section 4.3: Testing the Role of Frequency in LVC Productivity.

4.2 Computational Approaches to LVC Productivity & Acceptability

As mentioned previously, research into automatic detection of MWEs generally and LVCs specifically is closely related to the productivity of such expressions, since detection by systems trained on previously seen data must learn to adapt to detection of novel, previously unseen expressions. The manner and extent to which these expressions vary, and how much variability must be accounted for by detection systems, depends upon their level of productivity. Thus, there has also been a fair amount of research in computational linguistics on the productivity of

LVCs, and the variety of combinations that can occur to create an acceptable LVC. The

following sections give an overview of this research.

4.2.1 Determining Acceptable Combinations of *Light Verb + Noun*

Some researchers have attempted to find combinatory diagnostics that would allow

computational systems to accept *Mary took a bath* and reject *Mary made/did a bath*. Davis and

Barrett (2001) look to semantic roles for establishing such diagnostics. The authors use an

inheritance hierarchy of semantic roles, where, for example, the Actor thematic role has a more

specified subtype of Agent, or volitional Actor. They hypothesize that the subject argument of a

light verb must bear a role that is the same or more general than the role assigned to that

argument by the noun complement. For example, *make* is analyzed in this work as taking an

Agent subject, and this accounts for why *make* LVCs cannot involve complements that project

unintentional Actors. Accordingly, *Kim made a drop of the book* is unacceptable due to the

unintentional nature of the Actor in *dropping* events; however, *The army made an airdrop of*

supplies is acceptable due to the intentional nature of *airdropping* events. The authors find that

although this combinatory diagnostic is somewhat helpful, it fails to provide constraints ruling

out many unacceptable LVCs such as *Sandy took a belief/claim*. Although not noted in the work,

this diagnostic would also rule out a variety of relatively frequent LVCs found in the PropBank

corpus, such as *They made an appearance* and *They made progress* (*appearance* and *progress*

are not traditionally thought to require Agentive subjects, as exemplified in *The sun appeared on*

the horizon and progressed across the sky). In fact, Butt and Geuder (2001) cite attributing

volitionality or intentionality to subjects of events that are not normally volitional as a cross-linguistically common feature of LVCs, contrary to the assumptions of Davis and Barrett.

Acknowledging the limitations with this research, the authors turn to aspectual properties to find other combinatory diagnostics (Barrett & Davis, 2003). In this research, Barrett and Davis hypothesize that the patterning between the Aktionsart of the support verb and the noun is not random — some features must be compatible for the two to combine. The authors investigate this hypothesis by compiling lists of support verbs and nominalizations in each of the four Aktionsart categories; these attested LVCs are then recombined to create combinations of all verbs and nouns, resulting in 3,049 combinations. The authors had these manually annotated as either "good" or "bad" LVCs, and examined the correlation of "good" LVCs with certain aspectual characteristics. Barrett and Davis found that the majority of "good" LVCs do involve combinations where the Aktionsart of the light verb and that of the noun are the same. However, they also found that accomplishment and achievement light verbs have equal probabilities of combining with either accomplishment or achievement nouns. The authors establish a statistically significant correlation between acceptability and the Aktionsart categories of the light verb and noun, and conclude that the most likely candidate noun following a support verb is that which is of the same Aktionsart category.

Although the authors hope that their results can be applied for creating an estimation algorithm to choose the best candidate combinations in a corpus, the utility of these results seems limited since their applicability relies upon having a corpus where the Aktionsart category of all words is retrievable. Even within their own work, the authors admit that assigning the Aktionsart category reliably can be difficult, and part of the reason for the interchangeability of

accomplishment and achievement light verbs and nouns may be due to mistakes in coding of the

Aktionsart category. Furthermore, the adequacy of the corpus used to establish these findings is

questionable, since the authors say that they find no examples of stative light verbs combining

with activity or accomplishment nouns. These are readily found in the PropBank corpus, for

example, *have a laugh.*

4.2.2 Developing Measures of Acceptability

Other research in computational linguistics has focused on developing measures of the (semi-)

productivity and acceptability of LVCs. Stevenson, Fazly and North (2004) seek to develop a

computational approach for both characterizing the set of complements that can occur with a

given light verb and quantifying their acceptability. The authors focus on the light verbs *take,*

give, and *make*, calculating PMI for instances of these light verbs in combination with the

indefinite determiner *a* and a noun that is form-identical to a verb: *LV a V* constructions. The

authors hypothesize that for "good" LVCs, they will see a much higher PMI for this pattern than

for other *LV [det] V* patterns since it is assumed that "good" LVCs generally involve indefinite

determiners while "bad" non-LVCs may involve a variety of other determiner types. (It should

be noted here that this linguistic assumption runs counter to the findings of the corpus study

presented in Chapter 3, which showed that reference tracking, involving the use of a wide variety

of determiner types within LVCs, may be one of the primary functions of LVCs as opposed to

their counterpart lexical verbs.) Stevenson, Fazly, and North develop a combined measure that

takes into account the PMI of the *LV a V* pattern and the *LV [det] V* pattern, and examine how

well this measure captures acceptability in comparison to acceptability established by human

ratings of the LVCs. The authors create a data set of potentially "good" and "bad" LVCs by combining the light verbs of interest with verbs that are form-identical to nouns found in Levin's (1993) verb classes. The use of the verb classes is motivated by the intuition that these classes could capture the kind of semantic similarity that unites observed "families" of similar LVCs. The statistics needed to calculate the acceptability measures of these LVCs are drawn from web searches.

Although the authors obtain reasonably good correlations between the acceptability measures and human acceptability ratings (using Spearman Rank Correlation), classes with lower numbers of "good" LVC combos have poor correlations. Unfortunately, with the exception of *take* combined with complements drawn from Levin's Nonvehicle Motion 51.4.2 class (with members such as *cruise, drive, paddle*), less than 50% of the combinations derived from other Levin classes received a medium or high acceptability rating. This demonstrates that simply extending combinations with all members of a Levin class is not a feasible option.

The research presented in North (2005) is quite similar to that described in the preceding paragraph; however, in addition to an acceptability measure like that of Stevenson, Fazly and North (2004), which North dubs the "LVC-PMI" measure, North develops two other measures of acceptability and compares how each performs in predicting acceptability as compared to human rankings. North also uses basic PMI as an informed baseline measure of acceptability. One additional measure tested is the "LVC-Prob" measure, which is a single measure composed of several corpus frequencies relating to the LVC, inspired by the intuition that higher-frequency words are more likely to serve as complements of LVCs, and that the probability of LVC-hood in part depends on how frequently a particular complement participates in LVCs in general. The

second measure, "LVC-Freq," is an intentionally simple measure that could be used with linguistically naïve search engines. This measure simply rates candidate constructions for acceptability based on how frequently they are attested in the corpus, which is the world wide web. This final measure is somewhat problematic due to noise and false positives in the data. North attempts to overcome this issue by assuming a certain amount of noise. His estimate is based upon the intuition that LVCs are more likely to be expressed without internal modifiers, so he assumes that the likelihood of seeing an LVC with N modifiers approaches zero as N increases. Although at a certain point this may be true, it was established in Chapter 3 of this work that LVCs occur with 1-2 modifiers on average. Thus, decreasing the likelihood of LVC-hood based on 1-2 modifiers may have led to problematic filtering in North's research.

North also relies on the possibility that Levin (1993) classes may be good sources for candidate LVCs since certain light verbs seem to combine with certain semantic classes of nouns. However, North also uses representative seed words drawn from these Levin classes to find WordNet groupings of semantically similar nouns. The author then gathers human ratings of acceptability for all combinations of the light verbs *take, give,* and *make* combining with zero-derived nominals from the Levin classes and nouns from the WordNet groupings. Like Stevenson, Fazly, and North (2004), North's human annotations reflect relatively poor acceptability of the majority of combinations drawn from these groupings. Overall, only 31% of all LVCs formed from Levin classes and 27% of all LVCs from the WordNet groupings are rated as either "fair" or "good."

After establishing human rankings of acceptability for the LVCs and drawing the requisite statistical information for the various acceptability measures from the web, North uses

Spearman Rank Correlations to score each computational measure against human rankings. He

finds that the best correlations for the combinations from Levin classes are achieved by the LVC-

Prob measure, averaging 49% agreement. He finds that both LVC-PMI and LVC-Prob average

49% agreement across the combinations from WordNet. The simple LVC-Freq measure

performs worse than the baseline, possibly due to noise in the data. North then examines the

correlations with coarser-grained acceptability ratings by placing finer-grained numerical ratings

into the bins "poor," "fair," and "good." The observed agreement then is much better, with

scores from 71%-81% agreement for both Levin and WordNet classes across the different

measures. This coarser-grained measure may be somewhat problematic given that the majority

of combinations are "poor." Nonetheless, North concludes that LVC-PMI is a slight

improvement over the baseline, while LVC-Prob is a marked improvement over both the PMI

baseline and LVC-PMI. North takes this as evidence that the semantics of the complement are

highly relevant to measuring acceptability.

Fazly, North and Stevenson (2005) extend and improve upon this bedrock of research by

proposing a statistical measure that incorporates syntactic properties and firstly places light verb

usages on a continuum of meaning from literal to figurative, while also placing them on a

continuum of acceptability. This measure uses evidence of syntactic flexibility of a potential

LVC to situate it on this continuum, relying on the assumption that syntactic flexibility reflects

greater analyzability and compositionality (Nunberg, Sag & Wasow 1994). Specifically, the

authors establish a measure of syntactic flexibility based on corpus counts of potential LVCs

participating in certain syntactic patterns, such as co-occurence with referential determiners and

quantifiers, realization in the passive form, and pronominalization of the noun complement (e.g.

which groan did Azin give?). The authors compare the frequency of appearance of a potential

LVC combo in these syntactic patterns, compared to the frequency of that combination's

appearance in what is deemed the typical LVC pattern: *Light verb + a + noun.* The motivating

assumption is that phrases that appear more often in a single canonical form, showing less

syntactic flexibility, will more likely be acceptable LVC combinations. The authors also use

what appears to be the LVC-Prob measure from North (2005).

Fazly, North and Stevenson again gather the statistics needed for the acceptability

measures from the world wide web, and gather the statistics on syntactic flexibility for the

compositionality measure from the BNC (which provides more detailed syntactic information

than the web). The computational measures for acceptability are compared to the same human

rankings established for Stevenson et al. (2004) and the computational measures for

compositionality are also compared to manual annotations of compositionality gathered for the

research. For both sets of human annotation, the authors note that the agreement rates between

annotators are quite low for *give* (0.34 agreement for compositionality ratings, 0.39 for

acceptability ratings) and better for *take* (0.70 agreement for compositionality ratings, 0.72

agreement for acceptability ratings). The authors find that the compositionality measure drawn

from the BNC corpus data has reasonably high correlations with the human ratings. One

limitation of this compositionality measure, however, is that it requires certain syntactic

information that may not be efficiently obtained. A simpler measure of compositionality that

was applied to the web data (lacking such syntactic information) showed relatively poor

improvement over the baseline. Finally, the authors find that this acceptability measure

incorporating additional syntactic features has a much higher correlation to human judgments

than that established in Stevenson, Fazly, and North (2004).

Overall, these studies make notable progress towards providing computational measures

of LVC productivity. However, these studies also demonstrate some of the difficulties of

determining how LVCs are extended. Firstly, these studies are all limited to a specific subtype of

LVCs, of the form *LV + a + noun*, which may be a small majority of LVCs in actual usage, as

demonstrated in Chapter 3. Secondly, this research demonstrates how difficult it is to gather

acceptability ratings of LVCs. North (2005) points out that relatively frequent LVCs from the

web were rated as "poor" by the annotators since just two annotators can often be unaware of

certain constructions. This reinforces the complicated relationship between frequency and

acceptability, and also brings up the importance of context. Certain LVCs can be deemed

unacceptable when devoid of context, but when situated in the appropriate context they are

acceptable. Thus, the setup of stimuli is very important when gathering acceptability judgments.

These issues are considered in the development of surveys examining acceptability or

naturalness of LVCs for this dissertation research.

4.3 Testing the Role of Frequency in LVC Productivity

We'll now turn to original work that tests the roles of frequency and acceptability in relation to

LVC productivity.

4.3.1 The Frequency Hypothesis

Given the similarity of the Spanish *becoming* constructions to English LVCs, the hypothesis

developed in Bybee and Eddington (2006) seems quite applicable in the case of English LVCs as

well. To test Bybee and Eddington's hypotheses concerning the roles of frequency and semantic

similarity, this hypothesis was adapted to English LVCs:

> **Frequency Hypothesis:** Speakers will find novel LVCs acceptable if they are
> semantically similar to an attested, highly frequent LVC.

A secondary expectation that follows from this hypothesis is that novel or very low-frequency

LVCs that are semantically similar to an attested high-frequency LVC will be more acceptable

than LVCs that are semantically similar to attested low-frequency LVCs, since, as postulated by

Bybee and Eddington, these low-frequency exemplars may not be robustly represented in

linguistic memory, instead serving as "fringe" members of a cluster of exemplars. In their view,

such peripheral members should not be the primary basis of comparison, as a highly frequent

central member would be, in determining if an unfamiliar construction should be added to the

cluster of similar exemplars. To test this hypothesis, surveys of acceptability were firstly

developed via piloting. The pilot is briefly described below.

4.3.2 Pilot Experiments

Overview of experimental design: Native English speakers were asked to give an acceptability

rating to novel LVCs that are semantically similar to either a high-frequency original or a low-

frequency original LVC.

Subjects: The subjects of a pilot study were 36 students of a University of Colorado Boulder introductory linguistics course, *Study of Words*.

Materials: The LVC annotations of the British National Corpus (BNC) (manually annotated in Fazly, 2007) were firstly used to establish which LVCs are very high-frequency in the English language. Having established frequencies of LVCs, the next step is to examine the nouns in the highest frequency expressions, and determine what nouns are semantically similar. To do this, the nouns in high-frequency expressions were looked up in WordNet, or the derivationally related verbs of these nouns were looked up in VerbNet. The frequency of these synonymous terms in the BNC was also extracted, so that any frequency effects stemming from this term could be accounted for. Two WordNet synset or sister terms were then selected at random from both high and low-frequency ranges, and two VerbNet terms were similarly selected, culminating in four potentially novel LVC combinations. For example, the LVC combination *take + step* was found to be one of the most frequent LVCs in the BNC. The WordNet synset member *measure* was then randomly selected so that one test combination will be *take + measure*. In this case, and in other rare cases, the term selected at random creates an existing collocation: *take measures*. In most cases, the resulting combinations can be quite unfamiliar; for example, *make + trial* derived from *make + offer,* or *take + observance* derived from *take + peep.* Context was then added to these combinations to ensure that the test sentences modeled natural language. The contextual elements were drawn either from a survey of the original LVC, or a survey of contextual elements that frequently co-occur with the new noun. Two sentences were developed with elements from each context type, giving 4 test sentences for each novel LVC. This resulted

in 80 test sentences total, and these were split into two surveys consisting of 40 sentences each. An effort was made to ensure that each survey tested the same LVC combinations in both types of context, only the specific sentence differed across the two surveys.

Procedure: The pilot survey was created in Qualtrics survey software, as such it was administered online (http://ucsas.qualtrics.com/SE/?SID=SV_0lKGVut7SgpkmAl). Native English speakers were asked to judge each of these sentences on a five point scale ranging from "doesn't sound like fluent English and I can't make much sense of it" (score of 1) to "sounds like perfect English" (score of 5). The final sets of 40 sentences within each survey were randomized, and each survey was offered to about 30 subjects. As the survey was optional, one survey was taken by 12 subjects, and the other was taken by 24 subjects.

Pilot Results: The statistical analysis performed included 5 linear regression models predicting the odds of each response (1-5), given whether the LVC was based on a high or low-frequency original, controlling for subject groups (based on which survey version they received) and the source of the sentence's contextual elements. Although none of the models were statistically significant, there were trends for LVCs based on low-frequency originals to be rated as "2" ($t(59) = -1.87$, $p = .07$), or "1" ($t(59) = -1.82$, $p = .07$). Also of note is the fact that the median rating for novel LVC sentences using the original LVC context was 2, while the median rating for sentences using the noun's contextual elements was 3. This demonstrates the importance of providing plausible context for each sentence, and shows that speakers are very much in tune with collocational patterns of the language. Using the noun's context draws upon the speaker's

138

familiarity with local collocational patterns, while using the original LVC's context sometimes leads to very unfamiliar collocational patterns between the noun and its surrounding elements. Overall, the average rating for heavy, control sentences was 4.65, median 5, while the average rating for LVCs was 2.73, median 2. Therefore, novel LVCs are generally strange and unacceptable to speakers. However, context could still play an important role here in affecting acceptability judgments. This finding fits with the findings of Stevenson, Fazly & North (2004) and North (2005), who also found that the majority of combinations constructed by holding the light verb constant and substituting semantically similar nouns were unacceptable to speakers.

Limitations & Improvements: One potential source of error in this study is the assumption that semantically similar, novel LVCs can be created using WordNet and VerbNet membership. Members of a WordNet synset or sisterhood relation can vary greatly in how semantically similar they are, and VerbNet class membership is problematically based upon syntactic, not necessarily semantic, similarity. To address this issue, human judgments of semantic similarity of class members was initially considered. However, such judgments will be somewhat time-consuming and may be prone to error given the difficulty in pinpointing semantic (especially in contrast to pragmatic) similarity. Furthermore, it may be of greater utility if a computational lexical resource could be used fruitfully to examine the importance of semantic similarity in LVC productivity. Because FrameNet (Fillmore, Johnson & Petruck, 2002) considers real-world domains of semantic similarity, FrameNet may be a quality resource for collecting semantically similar nouns, so this resource was decided upon for future use.

It is clear from the pilot results that the contextual elements, and whether the new noun or original LVC's contextual elements are used, is very important and can strongly affect judgments of acceptability. Although great effort was made to ensure "natural" contextual elements in the test sentences, this is inevitably a source of error and an extremely time-consuming process. Thus, a new process for sentence creation was required (to be discussed in the next section). An unrelated, but also plaguing issue with the random selection of nouns within a frequency band was that it sometimes resulted in the creation of LVCs that were not novel. In fact, knowing whether or not a combination is truly novel would require intractable searches for the combination. Given both of these issues, the full study used extremely low-frequency LVC combinations found in the English Gigaword corpus (Graff & Cieri, 2003) of approximately 1,756,504,000 words. Undoubtedly, these combinations will be novel to some if not many speakers, but use of extremely low-frequency combinations allows for the use of attested contextual elements from a Gigaword usage.

4.3.3 Refinement of Hypothesis, Predictions

Although Bybee & Eddington were able to test "novel" *verb + adjective* combinations unattested in the corpus, our piloting (and other research, e.g. North, 2005) demonstrated that context strongly affects acceptability, and "novel" combinations cannot be presented to speakers in the original, sentential context natural to the utterance. Therefore, the stimuli of focus for the full testing were extremely low-frequency combinations attested in Gigaword. The Frequency Hypothesis was slightly refined to reflect this:

> **Frequency Hypothesis:** Speakers will find novel **or very low-frequency** LVCs acceptable if they are semantically similar to an attested, highly frequent LVC.

According to the Frequency Hypothesis, it is expected that the acceptability ratings of the very low-frequency LVCs that are semantically similar to high-frequency exemplars will be higher than the ratings for the very low-frequency LVCs that are semantically similar to low-frequency exemplars. Thus, the token frequency of the attested exemplar LVC is the main effect of interest in this study.

Furthermore, given the noted importance of context in the first pilot study, as well as the findings of the corpus study in Chapter 2, it is expected that LVCs involved in long sentences will be more acceptable than those involved in short sentences. Therefore, the full study included stimuli presenting the same very low-frequency test LVC in both a short and long sentence, and this was included as a variable in the statistical analysis. Additionally, an identifier for each unique stimulus sentence was included as a variable so that random differences in each sentence that may affect acceptability were also controlled for. Similarly, since different subjects certainly have distinct linguistic backgrounds and therefore distinct judgments, an identifier for each subject was included as a variable to be controlled for.

In Section 4.1.2, linguistic approaches to categorization and productivity were discussed. This research showed that type frequency, in addition to token frequency, may play a role in the extensibility of constructions. Specifically, a greater range of unique types can lead to greater variability of a category of acceptable constructions, and therefore a greater tolerance for new constructions. Similarly, the number of semantically similar unique types within a family or cluster of LVCs can reflect the coverage of a particular semantic space, which also affects extensibility as discussed in 4.1. Thus, the number of unique LVC combinations, or types, within

a semantically similar family was also noted. This information was included as a variable in the

statistical analysis so that it could be controlled for. Additionally, the token frequencies of each

unique LVC combination within a family may affect acceptability, so a variable that captures the

distribution of all combinations within a family across frequency bands was also included (this

will be described in greater detail in the Methods section to follow).

Finally, past research (Stevenson, Fazly & North, 2004; North, 2005) and piloting also

showed that different light verbs vary in their overall acceptability. Thus, the identity of the light

verb itself, in this case *give, have, make,* and *take*, was also included as a variable in the analysis.

4.3.4 Overview of experimental design

Native English speakers were asked to give an acceptability rating to very low-frequency LVCs

(occurring 20-50 times in Gigaword) that are semantically similar to either a high-frequency

exemplar (occurring 2,000-3,000 times in Gigaword) or a low-frequency exemplar LVC

(occurring 100-200 times in Gigaword).

4.3.5 Subjects

125 participants were paid $1.00 to complete the survey via Amazon's Mechanical Turk.[11] The

worker requirements were that they had been approved for 95% or more of their previous jobs

(Human Intelligence Tasks or "HITS"), and that the worker was located in the United States.

This is verified via IP address and helps to ensure that the participants are English speakers. The

quality of judgments from "Turkers" has been demonstrated in other studies involving judgments

---

[11] https://www.mturk.com

of acceptability of sentences involving island violations (Sprouse, 2011). The large number of

participants is needed to capture a significant difference in acceptability based on frequency, and

also ensures that there will be an adequate number of participants despite the fact that, on

average, about 10% of Turkers provide what seem to be random responses that must be thrown

out (Sprouse 2011; Bermel & Knittl, 2012a,b).

The keywords advertising this HIT expressed a desire for native English speakers, but

this was not required for participation (in the Mechanical Turk community, limiting HITs to

speakers of a particular language may encourage lying by the workers (Sprouse, 2011)). Within

the survey, participants were asked if English was their first and/or primary language. All

participants responded "yes" to this question, thus, no participants were eliminated for being a

non-English native speaker.

Subjects were weeded out by examining the responses to control stimuli of two types.

The first type were stimuli that were expected to be highly acceptable. These included heavy,

compositional usages of the same light verbs. These usages were selected based on frequent

collocations for the heavy sense, and the sentences were taken directly from Gigaword but

simplified by eliminating adjunct clauses (e.g. *Most students* **have computers** *in their homes*).

The control stimuli expected to be acceptable also included Gigaword usages of the highly

frequent exemplar LVCs (e.g. *He* **has a deep understanding** *of the problem and is trusted*

internationally). One sentence of each of these types was included for each light verb of interest.

The second type of control stimuli were sentences that were expected to be highly unacceptable.

These were the most poorly rated sentences drawn from the pilot experiment (e.g. *Pat* **made an**

administration *towards the goal*). Again, one decidedly bad LVC involving each light verb of

interest was included. Subjects who rated the first type of stimuli as highly unacceptable (from

0-10 on a scale of 0-100, where 0 is "odd" and 100 is "perfectly fine") were eliminated from the

study. Similarly, subjects who rated the second type of stimuli as highly acceptable (from

90-100 on the scale) were also eliminated. This resulted in the exclusion of 13 subjects, who

either were not paying adequate attention to the task or had highly unusual grammatical

judgments. Thus, the total number of subjects used in the study was 112.

4.3.6 Materials

The starting point for stimuli selection was a spreadsheet of potential light verbs (*do, give, have,*

take, make) and their noun collocates, drawn from the Gigaword corpus. For each collocation,

the frequency in Gigaword was listed.[12] The combinations in this spreadsheet were then filtered,

to include only combinations with a noun listed in WordNet as having a "type" (according to the

lexical file information) that could be eventive or stative. Specifically, nouns of the following

types given in Table 4.3.6, previously discussed in Chapter 2:

Lexical File Noun Type	Nouns denoting…
noun.act	acts or actions
noun.cognition	cognitive processes
noun.communication	communicative processes
noun.event	natural events
noun.feeling	feelings and emotions
noun.location	spatial position
noun.motive	goals
noun.phenomenon	natural phenomena

---

[12] Thanks to Afsaneh Fazly, personal communication

Lexical File Noun Type	Nouns denoting…
noun.possession	possession and transfer of possession
noun.process	natural processes
noun.relation	relations between things
noun.state	stable states of affairs

Table 4.3.6: WordNet noun types hypothesized to correspond to potentially eventive or stative nouns that may be involved in LVCs.

Next, a Java program was developed to take an LVC combination as input (e.g. *give_look*), and output all Gigaword combinations involving a noun sharing a FrameNet frame with the input noun, along with each combination's frequency in Gigaword.[13] Specifically, the output of this program is a spreadsheet displaying what can be thought of as "families" of semantically similar LVCs, and their frequency distributions in Gigaword. Families with two types of frequency distributions were selected: first, those that include a high-frequency LVC (2,000-3,000 instances in Gigaword), and a very low-frequency LVC (20-50 instances in Gigaword), and second, families that include a low-frequency LVCs (100-200 instances in Gigaword) and a very low-frequency LVC. After further analysis of the token frequencies of each LVC in a selected family (which will be discussed in greater detail in the Methods section), very low-frequency test LVCs were selected from the families. Then, stimuli sentences involving these LVCs were collected from Gigaword using regular expressions and Grep searches.[14]

---

[13] Thanks to A. Vichanthangal Prathivadhibayan, Computer Science PhD at CU Boulder, for collaborating in the creation of this program.

[14] Thanks to Alexis Raykhel, Linguistics Masters student at CU Boulder, for assisting in gathering potential LVC sentences from Gigaword.

4.3.7 Methods

In this experiment, FrameNet is used to determine what LVCs are "semantically similar." Where nouns combining with the same light verb share a FrameNet frame, they are considered semantically similar. In order to test the Frequency Hypothesis, we need to compare the acceptability of (1) very low-frequency LVCs that are semantically similar to high-frequency LVCs, to the acceptability of (2) very low-frequency LVCs that are semantically similar to low-frequency LVCs. The Frequency Hypothesis will be supported if the acceptability of (1) is greater than that of (2).

One challenge to testing this hypothesis is summed up by Zipf's law (1949): word-frequencies have a heavy-tailed distribution (i.e. there are a lot of infrequent types and few high-frequency types). This is also true of families of semantically similar LVCs: commonly, a few very high-frequency LVCs exist, with a long tail of many different low and very low-frequency LVCs that are semantically similar in the sense that the nouns share a FrameNet frame. In order to cleanly test the Frequency Hypothesis, one would ideally want to test the acceptability of very low-frequency LVCs that are solely related to either a high-frequency LVC or a low-frequency LVC. If this is not the case, then one cannot be certain that the acceptability is not influenced by the existence of other lower or higher-frequency LVCs. Thus, the families selected attempt to achieve this ideal scenario despite Zipf's law: families with the largest percentage of tokens in the high-frequency band are considered "high-frequency families" and families with the largest percentage of tokens in the low-frequency band are considered "low-frequency families." Tables exemplifying one low-frequency family and one high-frequency family are given below.

LVC	Frequency
make realization	5
make inference - very low-frequency test LVC	20
make deduction	24
make guess - low-frequency exemplar LVC	102
make conclusion	165

Table 4.3.7-1: A low-frequency "family" of LVCs detected in Gigaword: most tokens fall into the low-frequency band 100-200 instances. The nouns of the LVCs in these families share a FrameNet frame.

LVC	Frequency
have dread - very low-frequency test LVC	24
have terror	33
have apprehension	107
have fear - high-frequency exemplar LVC	2342

Table 4.3.7-2: A high-frequency "family" LVCs detected in Gigaword: most tokens fall into the high-frequency band 2,000-3,000 instances.

In the High-frequency Family, 93% of the tokens in the family (2,342 out of 2,506 total tokens) are *have fear* tokens falling into the high-frequency band. In the Low-frequency Family, 86% of the tokens in the family (267 out of 311 total tokens) are either *make guess* or *make conclusion* tokens, falling into the low-frequency band. Notably, it is only the presence of one high-frequency expression, *have fear,* that puts the second family into the high-frequency band.

To account for the potential effects of type frequency, the number of unique LVC combinations within a family was also noted (e.g. 5 unique types for the family exemplified in Table 4.3.7-1, and 4 types in the family in Table 4.3.7-2). Family sizes ranged from 3 members to 32 members, but the average number of members across both the high-frequency and low-

frequency families was the same: 11 members. Each family was also marked with a group name reflecting the distributions of each type across the frequency bands in Gigaword:

Short Tail: the other unique LVC combinations outside of the family's dominant token frequency band are all higher-frequency

Split Tail: the other unique LVC combinations outside of the family's dominant token frequency band are split—some are higher-frequency and some are lower-frequency

Long Tail: the other unique LVC combinations outside of the family's dominant token frequency band are all lower-frequency

No cases of Short Tail families were encountered, in accordance with Zipf's law. The inclusion of this group information allows for the frequency distributions of other unique LVC types to be accounted for in statistical analysis. Charts exemplifying a Split Tail and Long Tail family are given below. The previous tables (4.3.7-1 and 4.3.7-2) exemplifying low and high-frequency families were both Long Tail families, since all types outside of the designated frequency band of the family were lower-frequency types.
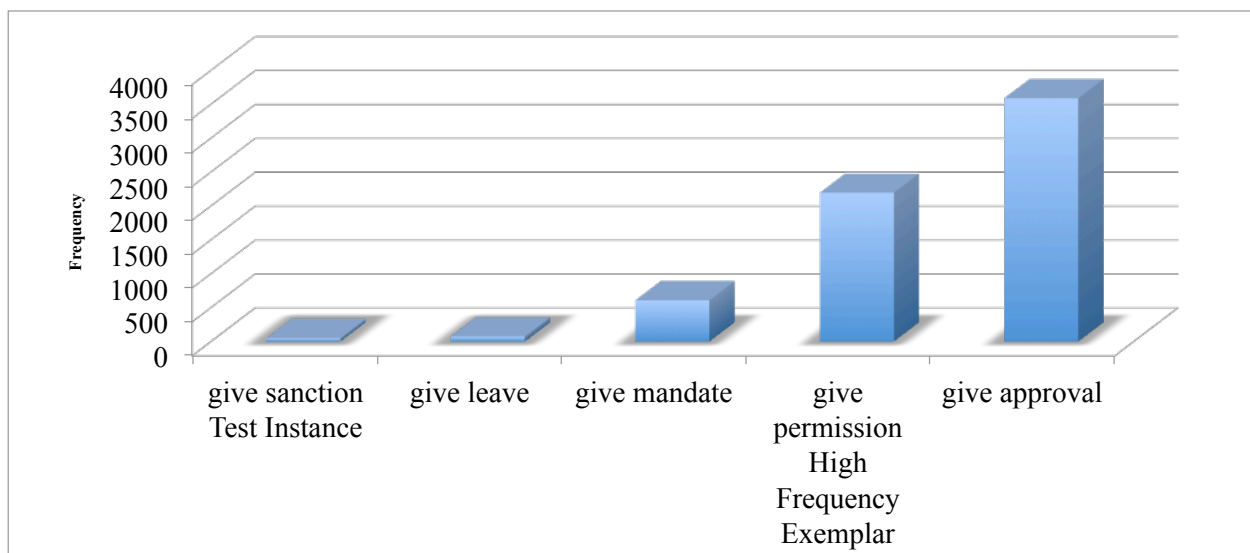

Chart 4.3.7-1: A "Split Tail" family — the high-frequency exemplar upon which the test instance is based has other types that are either higher or lower-frequency.
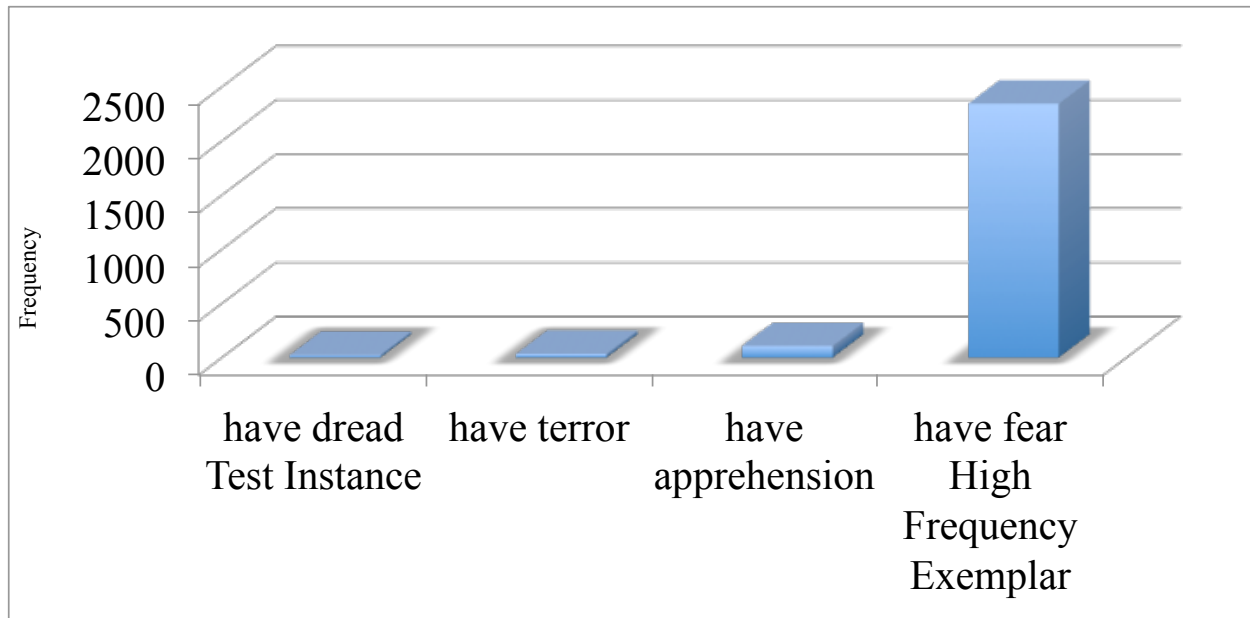
Chart 4.3.7-2: A "Long Tail" family — the high-frequency exemplar upon which the test instance is based has only lower-frequency types in the family.

For the twelve resulting high-frequency families, seven families were Long Tail families and five were Split Tail families. For the twelve low-frequency families, four families were Long Tail families and eight were Split Tail families. Unfortunately, there were not enough options for stimuli to balance this factor across the high and low-frequency families; nonetheless, it is included as a control variable in the analysis.

For each of the light verbs *give, have, make, take*, three high-frequency and three low-frequency families are selected. One very low-frequency LVC is selected from each of these families, and it is considered semantically similar to existing high-frequency LVCs if its family is high-frequency, and it is considered semantically similar to low-frequency LVCs if its family is low-frequency. This process resulted in 24 very low-frequency LVCs to be tested for acceptability — 12 that are semantically similar to an existing high-frequency LVC and 12 that are semantically similar to an existing low-frequency LVC.

To ensure that the exemplar and test pairs were equally semantically similar across all cases of high-frequency exemplars and low-frequency exemplars, LSA similarity was calculated for each exemplar and test noun.[15] LSA assigns a similarity score between -1 and 1, where a score closer to 1 indicates greater similarity. On average, LSA similarity between exemplar and test nouns for high-frequency exemplars was 0.22 and 0.20 for low-frequency exemplars. This shows that the semantic similarity between the exemplars and test nouns is quite comparable, and greater semantic similarity in one group or another will not affect acceptability judgments.

For each of the 24 very low-frequency LVCs, sentences containing these combinations in light usages were collected from Gigaword. Two sentences were collected for each LVC: a simple sentence and a full-context sentence. The simple sentence uses the existing Gigaword sentence, but excludes certain portions of the sentence. None of the content of the sentence is changed, but preceding or following adverbial phrases are removed to ensure a short, simple sentence. The full-context sentence is a sentence that is taken without any changes from Gigaword, and is intended to give rich contextual detail surrounding the event expressed by the LVC. There are several reasons why including this manipulation may reveal important information about LVC acceptability. Firstly, it was seen in the corpus study that the vast majority of sentences with LVCs contain modifiers of various sorts, as well as relative clauses and coordination. In general, it could be said that LVCs tend to occur in rather complex sentences containing descriptive elements. Thus, it is possible that the acceptability of LVCs in short, simple sentences will be lower because speakers are accustomed to encountering LVCs in longer, more descriptive and complex sentences. Similarly, sentence length may affect

---

[15] http://lsa.colorado.edu: a pairwise comparison was completed for each test noun and exemplar noun.

acceptability in relation to the Gricean Maxims of Quantity (give as much information as is needed, and no more) and Manner (be clear and brief, avoiding obscurity and ambiguity). In general, the use of LVCs versus lexical verbs should be in keeping with these maxims. Given these maxims, it may not be clear to speakers why one would elect to use a longer LVC to express a concept in a short sentence rather than a short, simple lexical verb. On the other hand, in longer, more detailed utterances, the use of an LVC may be more acceptable. Finally, it may be that longer sentences facilitate a certain amount of bootstrapping, such that speakers are able to better understand the meaning of an unfamiliar LVC in a longer sentence than in a shorter sentence.

Thus, a total of 48 sentences exemplifying the 24 very low-frequency LVCs are collected as stimuli in a Qualtrics survey. To serve as a control and better understand the acceptability of LVCs generally, sentences involving the exemplar LVCs that were used as the basis of comparison for the semantically similar very low-frequency LVC were also collected (for example, a sentence with the high-frequency LVC *give_permission*). One sentence was collected for each of the four light verbs, reflecting both high and low-frequency basis LVCs, for a total of eight control sentences. Finally, four control sentences involving the same verbs in their heavy usages are added, and four sentences involving *make* and *take* in a decidedly "bad" LVC (as determined in the pilot surveys) are also included. This results in a survey consisting of 64 sentences.

4.3.8 Procedure

Mechanical Turk was used to link participants to the Qualtrics survey site and pay them. These sentences were presented in random order to participants, and participants were asked to judge the acceptability of each sentence on a scale from 0-100, where 0 is "sounds odd" and 100 is "sounds perfectly fine." The scale itself is not seen by the participant, only the labels. Presenting the scale in this way allows for the ratings to be understood as a continuous scale rather than the Likert-type scale used in Bybee and Eddington (2006) and the pilot study, which are in fact categorical scales. There are several drawbacks to categorical scales, investigated in Treiblmaier and Filzmoser (2009): the data may not be normally distributed, and this will be exacerbated if researchers assign labels to individual categories or points along the scale (there is not necessarily an equal distance between, for example, 1-2 or "poor" to "less poor" on a 1-5 scale of acceptability), the data only allows for statistical techniques that do not rely on the arithmetic mean, the labels tend to influence subjects' responses, information may be lost due to limited resolution of the categories, and by constraining the range, the investigator may again influence response behavior. Continuous scales on the other hand, generate interval-scaled data and avoid the cognitive effort of matching semantic statements with interval points along the scale. Furthermore, Treiblmaier and Filzmoser note that the default position of the slider in such surveys can affect users as many may maintain the default position. Thus, in this survey, users were required to click on the scale in order for the slider bar to appear, and this action was needed before respondents could move on to the next question. A sample question is shown in
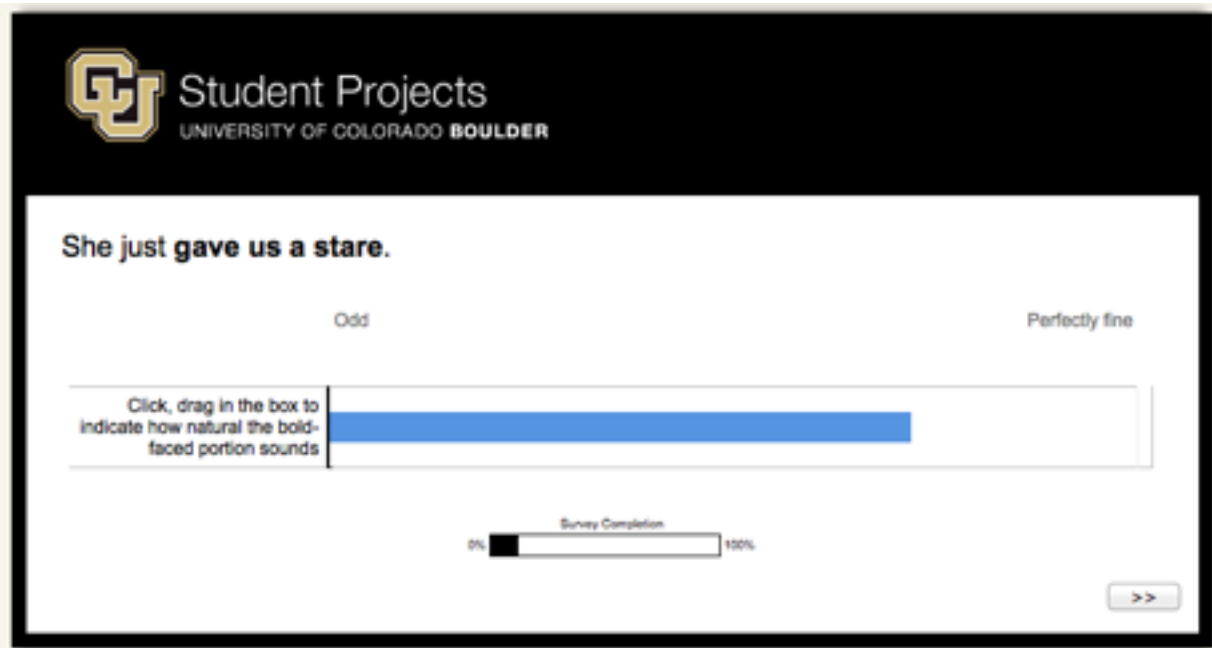
152

4.3.8.[16]

Figure 4.3.8: Example test instance from the Qualtrics survey administered via Mechanical Turk.

The survey was distributed to 125 participants via Amazon's Mechanical Turk, which allows for

efficient and affordable crowdsourcing of such judgments.

4.3.9 Results

After data collection, R (R Core Team, 2012) and *lme4* (Bates, Maechler & Bolker, 2012) were

used to perform a linear mixed effects analysis of the relationship between the subject's

naturalness rating of the very low-frequency LVCs and the frequency band of the semantically

similar exemplar LVC (low-frequency, 100-200 instances in Gigaword or high-frequency,

2,000-3,000 instances in Gigaword).[17] As fixed effects, the following variables were entered: the

---

[16] The full survey can be viewed here (simply enter any number as the Mechanical Turk Worker ID if you wish to proceed through some questions): https://cuboulder.qualtrics.com/SE/?SID=SV_b1KZBFNFtQNwvL7

[17] Full results can be found here: https://docs.google.com/spreadsheets/d/1GHO4KTSvwkQwguGCqc7BgU0S9liqV9mgNfmopGOuNQY/edit?usp=sharing

dominant token frequency band of the family (high or low), the verb type (*give, have, make, take*), the number of unique LVC types that were semantically related via FrameNet membership in a "family," the Long Tail or Split Tail label, and the long or short context type. As random effects, there were intercepts for subjects and items (stimuli sentences), as well as by-subject and by-item random slopes for the effect of the frequency band of the semantically similar LVC. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question.

The frequency band of the semantically similar exemplar LVC does significantly influence the naturalness rating for the very low-frequency LVC, but not in the manner that was expected — very low-frequency LVCs that are semantically similar to a low-frequency LVC (100-200 instances) are 8.56 points (on a scale from 0-100, +/- 3.50 standard errors) more acceptable than very low-frequency LVCs that are semantically similar to a high-frequency exemplar (2,000-3,000 instances) ($X^2(1)=5.64$, p=0.01751). This is a fascinating result that evidences the importance of statistical preemption or "blocking" in extending constructions (Goldberg 2006, Suttle & Goldberg, 2011), which will be described in greater detail in the discussion.

The distribution of all unique LVC types in the family across frequency bands (i.e. the Long Tail or Split Tail family type) is also a significant predictor of acceptability rating. Recall that for some very low-frequency LVCs tested, the other types outside of the majority frequency band of the family are split between being higher-frequency or lower-frequency than the majority frequency band of the family (Split Tail family type). For other very low-frequency LVCs tested, the other types outside of the frequency band of the family are all lower-frequency than the

majority frequency band of the family (Long Tail family type). Very low-frequency LVCs from Long Tail families tend to be about 8.02 points more acceptable (on a scale from 0-100, +/- 3.53 (standard errors)) than those LVCs that are from Split Tail families ($X^2(1)$=4.84, p=0.02775).

It was also found that the light verb itself is a significant predictor of acceptability rating ($X^2(1)$=10.30, p=0.01616): the very low-frequency LVCs involving "give" tend to be the most acceptable, those involving "make" are the second most acceptable, "have" is the third most acceptable, while those involving "take" are the least acceptable. This result is consistent with the findings of North (2005), who found that *give* LVC combinations garnered the greatest number of "fair" or "good" ratings. North found that *take* and *make* LVC combinations were generally less acceptable than *give* LVCs, but they were not significantly different in their overall acceptability. He did not include *have* LVCs for examination.

Although it was predicted that LVCs in longer sentences would be more acceptable than those shown in short, simple sentences, the context (short or long) was not a significant predictor of acceptability (this point is discussed in greater detail Section 4.5). Furthermore, the number of unique LVC types within a family was not a significant predictor of acceptability. Finally, the main variable of interest (high vs. low token frequency of the attested exemplar) did not interact with any of the other variables.

4.4 Discussion of Results

These results are best understood through the lens of statistical preemption or semantic bleaching. Each of these processes will be discussed in turn in the sections to follow, along with related work in cognitive science and psycholinguistics.

4.4.1 Statistical Preemption

As mentioned previously, the fact that the very low-frequency test LVCs were found to be

significantly more acceptable if they were semantically similar to a low-frequency exemplar, as

opposed to a high-frequency exemplar, is consistent with a process of statistical preemption or

"blocking" of a lower-frequency, marked alternative by a higher-frequency, unmarked variant.

The process of statistical preemption has been touched on briefly in the discussion of the

relationship between frequency and acceptability in Section 4.1, but will be examined in light of

these findings here. In the process of statistical preemption, speakers implicitly infer from

consistently hearing a formulation, B, in a context where one might have heard a semantically

related alternative formulation, A , that B is the appropriate formulation and A is not appropriate.

A simple example of preemption occurs in morphology, where, for example, *did* preempts *do-ed*

(Suttle & Goldberg 2011). In the case of LVCs, speakers may infer that a high-frequency LVC is

appropriate; for example *give a look*, as in *She gave them a dirty look.* Thus, they may also infer

that a semantically similar, rare LVC is not appropriate; for example, *give a stare,* as in *He gave*

me a stare that didn't look as if he was amused (actual test item in survey, from Gigaword). In

simple terms, speakers are so accustomed to hearing certain high-frequency LVCs in a given

context that it simply sounds odd when a speaker uses an alternative expression where the high-

frequency expression seems viable.

On the other hand, if the very low-frequency LVC is semantically similar to a

low-frequency LVC, then statistical preemption does not occur, but it is likely that even

low-frequency LVCs occur often enough that speakers can extend this pattern by analogy.

Indeed, Bybee (2010) finds evidence that speakers retain some details of each token of linguistic experience, and Bermel & Knittl (2012) show that the relationship between frequency and acceptability is not straightforward, in the sense that low-frequency does not entail low acceptability to speakers.

Further evidence for statistical preemption is provided by the fact that Long Tail families were significantly more acceptable than Split Tail families. For the reader's convenience, the graphics illustrating each type of family are repeated here. In the Long Tail family, all other types outside the majority frequency band (where the majority of tokens fall – in the 2,000-3,000 band) are lower-frequency:
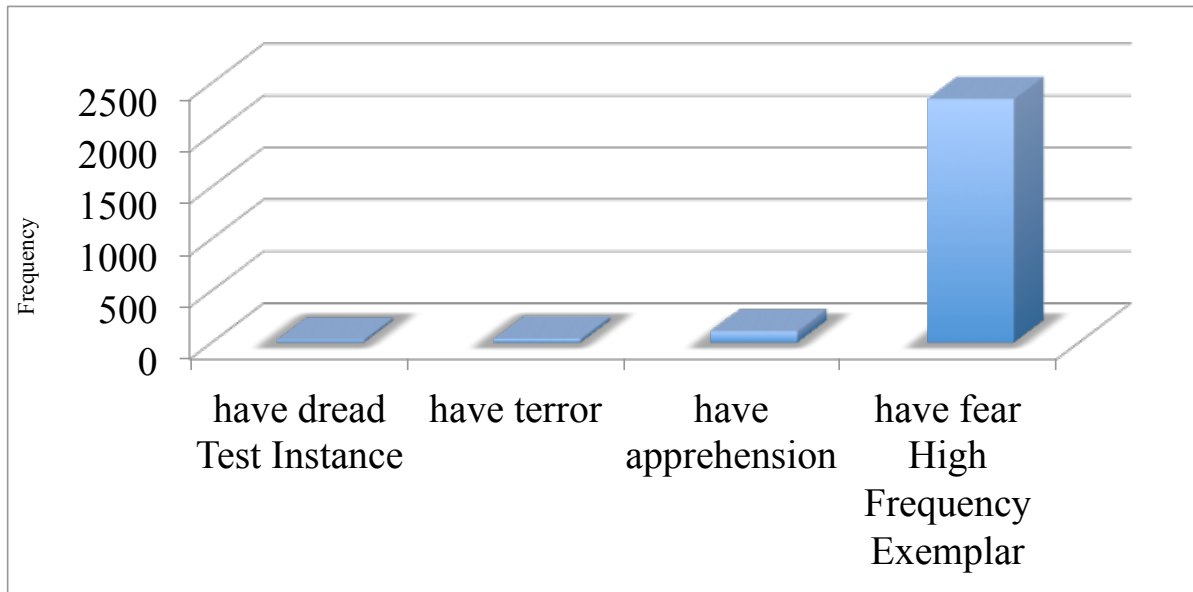

Chart 4.4.1-1: Long Tail Family — all other types are lower-frequency than the basis LVC, offering only one variant for preemption

In the above type of distribution, there is only one higher-frequency LVC that may preempt the usage of a very low-frequency test instance.

Now consider the other type of distribution, where the types are split – some types are lower-frequency and some are higher-frequency than the majority frequency band:
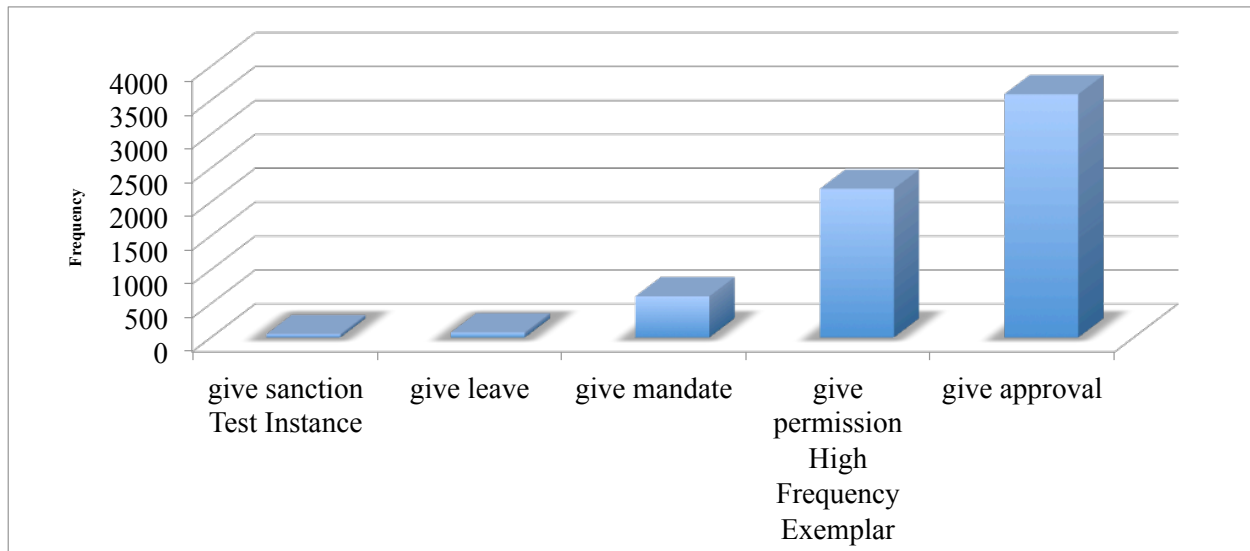

Chart 4.4.1-2: Split Tail Family — one type is higher-frequency than basis LVC, offering two variants for preemption.

In this family, *give_approval* was outside of the frequency bands of interest since it has a frequency in the 3000-4000 range. Nonetheless, in this Split Tail type of family, it's possible that both *give_permission* and *give_approval* potentially preempt the use of the very low-frequency test LVC. Thus, the significance of the distribution of other LVC types also shows that where there is a greater possibility of preemption, the very low-frequency test LVC will be found to be less natural. In these cases, there are two semantically similar, alternative formulations that are more frequent, alternative formulations instead of just one, increasing the likelihood of preemption by one (or both) of these forms.

These findings do not necessarily disprove the hypothesis that very low-frequency or novel constructions will be acceptable if they are semantically similar to a higher-

frequency exemplar (Bybee, 2010; Bybee & Eddington, 2006). Instead, it shows that perhaps "higher-frequency" simply doesn't have to be very high-frequency, and at a certain point, the high-frequency of an existing exemplar may actually block extension rather than facilitate it. Bybee and Eddington (2006) use a corpus of about 2 million words (as compared to Gigaword's 1,756,504,000 words). Furthermore, what they consider a "higher-frequency exemplar" in the Spanish *becoming* constructions data occurs anywhere between 4 (in the case of *quedarse gusto* – 'become pleased') and 29 times (in the case of *quedarse quieto* – 'become calm, still'). Thus, this research uses a very different corpus and very different frequency bands that are considered "high-frequency." Bybee and Eddington's notion of "high-frequency" is truly a relative term. Thus, this study adds to the research of Bybee and Eddington, demonstrating that there is a point at which statistical preemption occurs, blocking instead of facilitating extension of a particular pattern of *light verb + semantically similar noun*.

4.4.2 Statistical Preemption: Related Work

Statistical preemption has been demonstrated in Goldberg's (2011) research on verbs that demonstrate the ditransitive/dative alternation and those that do not. Goldberg notes that semantically similar verbs may be distinct in their patterning of realization in these forms: *She told her the news,* but *?She explained her the news*. Although Goldberg points out that the role of statistical preemption when considering phrasal forms is potentially unclear, given that phrasal forms are never semantically or pragmatically identical to another phrasal form, she concludes that semantic considerations do not

provide an explanation of the ill-formedness of *explain* in the ditransitive since the

meaning is perfectly semantically compatible with the ditransitive construction.

Goldberg develops a measure of the likelihood of statistical preemption. She proposes

that the probability of Construction B (CxB) statistically preempting Construction A

(CxA) for a particular verb, V can be determined by:

- P(CxB | a discourse context in which the learner might expect to hear CxA(with V), i.e.
- P(CxB | a discourse context at least as suitable for CxA and V

For example, there is a high probability for P (dative | a discourse context at least as suitable for

the ditransitive and *explain*), thereby indicating that *explain* doesn't readily occur in ditransitive.

Goldberg admits that determining what is "a discourse context at least as suitable for CxA and

V" can be difficult in practice. She simplifies this notion by simply finding the total number of

ditransitive and dative usages of a verb; in other words:

- P(dative | V and (dative or ditransitive))

However, Goldberg specifies that the semantics and information structure of the ditransitive must

be satisfied in these contexts for the ditransitive to be an "equally suitable" alternative. To

ensure that this is the case, Goldberg limits her study to syntactic contexts where it is known that

the ditransitive is preferred: when the Recipient is pronominal and the Theme is not: *She told me

the news* vs. ?*She told the news to me.* Goldberg collects the relevant counts for a set of

alternating and non-alternating verbs. She finds alternating verbs have a resulting probability of

the dative in these contexts that is .04 on average, whereas non-alternating verbs like *explain*

have a probability of .83 on average. Goldberg takes this as corpus evidence of the process of

statistical preemption, and further concludes that statistical preemption is an important factor in

the learning of arbitrary distributional restrictions, like the restrictions governing LVC combinations. Goldberg's study does seem somewhat somewhat problematic, however: she does not address the fact that, for example, *tell,* can occur in the transitive with a Recipient object (*I told her*), while *explain* can only occur in the transitive with a Theme object (*I explained it*). The frequency of these realizations may also play a role in the acceptability of the dative.

Although the connection has not been explicitly made as far as this author is aware, the process of statistical preemption is also consistent with research in cognitive science and psycholinguistics. Research using electroencephalogram (EEG) and event-related potentials (ERPs) has demonstrated the important effects of frequency-based expectations in sentence parsing. Such studies are able to examine electrical activity in the brain plotted over time. Early studies established that semantically anomalous words in a sentence elicit a negative shift around 400 milliseconds after the onset of the anomalous word, called an "N400" effect (Kutas & Hillyard, 1984). Syntactic anomalies, in contrast, elicit a positive shift at 600 milliseconds, called the "P600" effect (Hagoort, Brown & Groothusen, 1993). In more recent studies, it was also found that implausible verb-argument combinations (e.g. *the hearty meal was devouring…*) can elicit the P600 effect although such instances do not involve an outright syntactic anomaly (Kim & Osterhout, 2005). Based on these findings, it seems plausible that processing for sentences with implausible verb-argument combinations can be dominated by "semantic attraction," for example, between *meal* and *devouring.* This attraction overwhelms syntactic cues causing the well-formed syntactic cues to appear ill-formed, and produce a P600 effect (Kim & Sikos, 2011). In other words, speakers have what is partially a frequency-based expectation that

meals are *devoured* and do not do the *devouring,* leading such sentences to be deemed syntactically ill-formed.

To further examine this phenomenon, called the "Semantic P600," Kim and Sikos (2011) explore the corollary that semantic attraction is greater where it is enhanced by syntactic plausibility. The authors therefore compare the effect of a sentence requiring only one modification to make it syntactically and semantically felicitous to those that require more than one modification. Specifically, *The hearty meal was devouring* requires only one change to shift it to the felicitous *The hearty meal was devoured,* while *The hearty meal would devour* requires multiple edits. Kim and Sikos hypothesize that the single edit repair sentence will show the P600 effect, because the semantic attraction is bolstered by syntactic plausibility, while the multiple-edit repair sentence will cause the N400 effect typical of semantic anomalies. The authors do indeed find that there is a P600 effect for the single-edit repair sentences, and no such effect is found for the multiple-repair sentences. Kim and Sikos in part attribute this finding to "pattern completion," wherein speakers complete recognized patterns of event representations, based in part upon the frequency with which speakers experience certain words associated with certain events, such as *meals* being *devoured.* It can therefore be concluded that the P600 effect can arise in cases where pattern completion conflicts with the observed sentence.

Husain, Vasisth and Srinivasan (2014) demonstrate that expectations associated with pattern completion are very strong in the case of Hindi LVCs, realized in the form of *noun + verb.* In this research, the authors firstly use a self-paced reading study to show that when a frequency-based expectation for an upcoming part of speech is dashed, parsing the rarer structure consumes more processing time than parsing the frequent structure. From the results of this

162

study, the authors develop the hypothesis that expectation-based effects will be stronger in cases where a stronger prediction is made for a particular verb, as opposed to sentences where the prediction is less strong (i.e. for just a verb, as opposed to a certain verb or a verb with certain properties). Husain, Vasisth and Srinivasan test this refined hypothesis by making use of the expectations involved with Hindi LVCs. The authors use a completion study (subjects complete a sentence with a sentence-final blank) to firstly demonstrate that when Hindi speakers are presented with certain nominals involved in Hindi LVCs, these cue strongly for a very specific verb. In the completion study, it was shown that subjects were able to correctly provide the verb for 86% of sentences involving LVCs, and only 18% of sentences involving heavy *noun + verb* constructions. The authors secondly compare reading times where distances between arguments and the verb are manipulated for sentences with and without nouns involved in LVCs. The experiment shows that as the distance increases for sentences with LVCs, reading times are faster, whereas as distance increases with heavy *noun + verb* constructions, reading times are slower. The authors attribute the speedup in the LVC condition to the increased expectation of seeing the already predicted, exact verb, due to pruning of alternative structures. This research shows that LVCs strongly cue frequency-based expectations and pattern completion.

Given Husain, Vasisth and Srinvasan's (2014) findings that the nouns of Hindi LVCs strongly cue for a particular light verb, it seems plausible that English speakers, too, strongly expect certain LVC combinations, and when these expectations are not met, the speaker's pattern completions are in conflict with the observed sentence, potentially generating a P600 effect.

4.4.3 Semantic Bleaching

Another plausible explanation for these results is that the more frequent, and more entrenched

LVCs are semantically bleached, and therefore not decomposed into the semantics of the

individual component words, as suggested by Hopper and Traugott (1993). Essentially, subjects

may not recognize the semantic relationship between the noun of a very low-frequency LVC and

that of a high-frequency LVC because the subjects are not decomposing the semantics of the

high-frequency LVC. Conversely, subjects do semantically decompose the low-frequency

exemplar LVCs, and are therefore able to see the analogical connection between these

constructions and their semantically similar, very low-frequency counterparts, resulting in higher

ratings for these LVCs.

4.4.4 Semantic Bleaching: Related Work

Sikos et al. (2009) have suggested that the processing of figurative language may be strongly

shaped by type and token frequency. The authors present a hypothesis wherein figurative

language is processed in two stages — first a literal interpretation is attempted, then rejected,

leading to a figurative interpretation. However, the authors note that in cases where the

figurative language is more entrenched, this process may be short-circuited, leading to direct

retrieval of the figurative interpretation. Although the work is preliminary, essentially the

authors propose that figurative language may sometimes be processed online and sometimes be

lexicalized, and that this distinction is determined by frequency: low-frequency instances may

require two-stage processing, while highly frequent instances may manifest a well worn path

through the memory, thus the figurative interpretation in these cases becomes automatic or lexicalized. Thus, it's plausible that there is a well-worn pathway in memory for highly frequent, entrenched LVCs and the semantically similar, very low-frequency variants simply seem like a "bad" version of entrenched expression, in part because the semantic similarity between the two is not noted since the semantics of the entrenched LVC are not decomposed.

However, other research in cognitive science and psycholinguistics exploring the processing of MWEs, including LVCs, does not support this interpretation. Cutting & Bock (1997) attempt to probe how idioms are stored in the mental lexicon, and conclude that idioms are not stored as unanalyzed chunks. In this study, speakers are briefly shown pairs of expressions, and after a short delay, the speakers are asked to produce one of the two expressions. The types of expression pairs are manipulated, such that in one experiment pairs of idioms with either the same meaning and same syntax, different meanings and same syntax, or different meanings with different syntax are presented. The authors then compared reaction times and errors in the speaker's production of a prompted idiom for each condition. They find that the reaction times are slowest in cases where the paired idioms are similar in both syntax and semantics, and errors such as producing *That's the way the cookie bounces* (a blend of *That's the way the ball bounces* and *That's the way the cookie crumbles*) are most frequent in all cases where the syntax is similar, regardless of semantics. The authors take this as evidence that speakers are sensitive to, and analyzing, the internal syntactic structure of idioms in addition to the figurative semantics. In a second experiment, speakers are presented with pairs of expressions sharing the same syntactic structure, either involving an idiom and a literal

phrase with the same meaning, an idiom and a literal phrase with a different meaning, two idioms with the same figurative meaning but different literal meanings, and two idioms with different figurative meanings and different literal meanings. In this experiment, the authors find that the conditions in which literal meanings overlap with figurative meanings result in more blend errors when speakers attempt to reproduce a phrase than the other conditions, despite syntactic similarity in all cases. Cutting and Bock interpret this finding as evidence that speakers are decomposing idioms and are simultaneously aware of the literal, compositional meaning as well as the figurative meaning. However, Cutting and Bock do not investigate whether or not such blend errors are affected by the frequency of the idioms presented. Perhaps more frequent, entrenched idioms would be less susceptible to such errors if, indeed, more frequent, entrenched idioms are more semantically bleached and lead to direct access of idiomatic meanings.

In related research, Briem et al. (2009) use an Magnetoencephalography (MEG) study to explore processing differences of verbs that can be used in either heavy or light usages, as compared to verbs that are unambiguously heavy. MEG is a technique that measures the magnetic fields generated by neuronal activity of the brain. The authors compare MEGs from three conditions: 1) both heavy and potential light verbs presented in isolation, 2) both heavy and potential light verbs with minimal context of a personal pronoun subject, and 3) only potential light verbs presented in a full context that disambiguated the heavy or light usage. The authors find that there is distinct cortical processing of potential light verbs, regardless of whether they are presented in a disambiguating context or not. Specifically, potential light verbs produce a smaller

activation, which the authors indicate is consistent with the interpretation that light verbs are perceived as underspecified in their semantics until the disambiguating noun object is presented. In other words, it seems that listeners "wait and see" before a particular sense is invoked for potential light verbs, but a sense is invoked nonetheless, indicating that the LVC is not stored as an unanalyzed lexical item.

Piñango, Mack & Jackendoff (2006) as well as Wittenberg & Piñango (2011) also find that there is evidence for independent processing, or decomposition of the verb and noun for LVCs in English and German respectively. These authors aim to explore whether LVCs are stored as pre-specified constructions in the mental lexicon, as would be predicted by Construction Grammar, or if the light verb sense is stored with heavy senses, and the light sense is triggered by a predicating noun complement. The authors then make the following possible predictions: according to the first interpretation, given the rather high-frequency of LVCs, one would expect lower processing costs and therefore faster processing of LVCs; according to the second interpretation, the semantics of the verb cannot be disambiguated toward the light sense until the noun is retrieved, leading to greater processing costs and slower processing of LVCs.

Both sets of research use a cross-modal lexical decision paradigm in which participants are asked to read sentences with either LVCs, non-light constructions with the same noun, or non-light constructions with the same verb. They are then asked to determine whether an unrelated word presented on the screen is a word of their native language (English or German). Reaction times are recorded when the word is presented either at the point that the verb is read, or 300 ms after the verb is read, approximately

167

where the noun is read. The motivation behind this paradigm is that processing costs will

be reflected in greater reaction times to the secondary task. Both sets of research also use

LVC stimuli that draw upon Kearns (2002) distinction between true light verbs and vague

action verbs (made on the basis of their ability to undergo certain syntactic operations,

such as the passive — vague action verbs passivize: *A speech was given by the president*).

Acknowledging that the very high-frequency of more idiomatic true light verbs may

affect processing time, the authors restrict their study to vague action verbs. In terms of

this dissertation research, vague action verbs are more likely to be considered semi-

LVCs, with arguments plausibly stemming from either the verb or noun because they are

more compositional and syntactically flexible than true LVCs.

Both sets of experiments also find that there is significantly greater reaction time

for LVCs when the lexical decision task is prompted at the noun position, indicating that

LVCs have a significantly higher processing cost, and supporting the interpretation

wherein speakers wait to disambiguate the semantics of the verb (i.e. determine whether

it is a heavy or light sense) until the noun has been retrieved. The authors also point out

that the ensuing process of determining how arguments are shared across the verb and

noun lead to greater processing cost as well.

While this body of research provides preliminary evidence for the processing of

some LVCs, it seems somewhat problematic that LVCs involving vague action verbs

were specifically selected for this work. Such LVCs, which are more likely to be semi-

LVCs, certainly involve additional processing because both the argument structure of the

noun and verb can be invoked and compete for disambiguating the construction.

Furthermore, the authors admit that many true LVCs, which tend to be more formulaic, syntactically fixed, and less compositional, may indeed be stored in the lexicon, and this is why vague action verbs were selected. Thus, this research is illuminating concerning one portion of LVCs, but does not necessarily indicate that all, or even most, LVCs (especially higher-frequency LVCs) are not stored in the lexicon as an unanalyzed chunk.

4.4.5 Conclusions

This research showed that very low-frequency LVCs that are semantically similar to high-frequency LVC exemplars are actually, somewhat surprisingly, less acceptable to native speakers than very low-frequency LVCs that are semantically similar to low-frequency LVC exemplars. This result could be explained through either statistical preemption or semantic bleaching. In the first account, the highly frequent exemplar LVC blocks the use of the semantically similar, very low-frequency LVC. Thus, when speakers judge the acceptability of a relatively unfamiliar LVC that is clearly quite similar to a very familiar construction, the unfamiliar LVC seems odd given that the familiar construction could have been used quite felicitously in the same context. In the second account, when speakers judge the acceptability of an unfamiliar LVC, they are comparing it to previously experienced constructions for similarity in order to decide whether or not the unfamiliar construction should be added to an existing family of similar constructions. However, the relationship of semantic similarity of the unfamiliar construction and the very high-frequency existing construction is not clear because the high-frequency construction has become entrenched to the point where speakers no longer analyze the individual components of the

expression; therefore, the analogy between the two is not recognized, and the unfamiliar construction is not categorized as a member of the existing cluster of exemplars.

Given the facts presented in the survey of related work, the first interpretation involving statistical preemption seems more likely. While the process of statistical preemption has been demonstrated quite convincingly in experimental settings (e.g. Boyd, Ackerman & Kutas, 2012), the process of semantic bleaching has not. In fact, although the process of semantic bleaching seems very plausible from a diachronic perspective, a growing body of research in psycholinguistics and cognitive science has demonstrated that speakers do recognize and analyze the individual lexical items within LVCs and idiomatic expressions (Cutting & Bock, 1997; Piñango, Mack & Jackendoff, 2006; Briem et al., 2009; Wittenberg & Piñango, 2011), contrary to semantic bleaching predictions. It should be noted, however, that these studies have not manipulated frequency to determine if the frequency of an idiom or LVC affects processing. Instead, existing studies have controlled for frequency. It would certainly be worthwhile to extend studies into the processing of LVCs to true LVCs, and to examine LVCs of varying frequencies. Until such studies are undertaken, current research supports an interpretation of these facts according to statistical preemption, as the process of semantic bleaching remains uncertain with regards to LVCs.

4.5 Limitations & Future Work

Despite taking great pains to ensure that the Frequency Hypothesis could be tested without interference from outside factors, a few potential sources of error that are very difficult to control for should be noted.

4.5.1 Context

The body of research into the relationship between frequency and acceptability has consistently

demonstrated the importance of context in judgments of acceptability (e.g. Bermel & Knittl,

2012b; Stevenson, Fazly & North, 2004; North, 2005). When testing the Frequency Hypothesis,

the importance of context was carefully considered in the creation of stimuli, and it was for this

reason that the actual sentential contexts surrounding the test LVCs was used. Given the findings

of Chapter 3 on LVC function, it was expected that LVCs situated in a longer, more informative

context may be considered more acceptable than those in a short, simple context (that may be

atypical of LVC usage more generally). Notably, the context type was not a significant predictor

of acceptability. This fact is perhaps more perplexing when we consider the potential

relationship between statistical preemption and context. Specifically, it's possible that certain

syntactic contexts cue for certain LVC combinations more strongly than others. For example, a

simple Google search for "a look that" returns about 99,400,000 results, and the search for "a

stare that" returns about 682,000 results. Thus, one can surmise that the noun *look,* and in turn

the LVC *give a look,* cues much more strongly for a *that* complement than the noun *stare* and the

LVC *give a stare.*

The strength of the correlation between a particular LVC and syntactic context depends

upon how often the LVC is realized in that syntactic context, compared with how often the LVC

is realized in other syntactic contexts. If the LVC is frequent in a variety of syntactic contexts,

then one can assume that a particular syntactic context may not cue as strongly for a particular

combination. If the LVC is consistently found in a single, or very few, syntactic contexts, then

we would expect the context to cue strongly for that particular LVC. In the latter case, if the LVC encountered is not the combination cued by the context, it will certainly seem "odd" and out of place. This is precisely what underlies statistical preemption: speakers implicitly infer from consistently hearing a formulation, B, in a context where one might have heard a semantically related alternative formulation, A , that B is the appropriate formulation and A is not appropriate. Thus, the strength of association between a particular context and a particular LVC combination may affect acceptability and preemption independently of the token frequency of the basis, exemplar LVC.

Intuitively, one might expect the longer, more informative contexts to provide the additional scaffolding that would cue for a particular LVC combination and plausibly lead to statistical preemption, but this was not supported by the statistical analysis. A closer examination of the stimuli; however, shows that the skeletal syntactic contexts (as far as what types of complements and prepositions follow an LVC) are generally quite similar across the two contexts, so this may explain why context was not a significant predictor of acceptability.

A preliminary followup study (somewhat like the completion study undertaken by Husain, Vasisth and Srinvasan (2014)) was undertaken, wherein the long context test sentences from the acceptability study were presented to 26 native English speakers, but the noun of the LVC was replaced with a blank.[18] The subjects were told that the study examined variability of expressions like *take a walk, make an offer, give a speech,* and *have a laugh;* therefore the subjects were already attuned to the concept of LVCs. They were then asked to examine 24

---

[18] The survey was created in Qualtrics (view here: https://cuboulder.qualtrics.com/SE/?
SID=SV_5alEafKD0MTDoSp) and subjects volunteered for the study via a link provided on Facebook
(www.facebook.com). A spreadsheet with the full list of responses is available here: https://drive.google.com/file/d/
0B-54bBqKapwXQTFQNUNrQ2xVWDQ/view?usp=sharing.

sentences and fill in the blank with the word that made the most sense to them, carefully

considering the context. They were presented with 12 sentences involving test LVCs related to

high-frequency exemplars and 12 sentences involving test LVCs similar to low-frequency

exemplars, in random order.

For sentences involving test LVCs related to high-frequency exemplars, subjects filled in

the blank with the exact noun of that high-frequency exemplar in 19% of the sentences. For

example, the very low frequency LVC *give a stare* was found in the following Gigaword

sentence:

99. He gave me a stare that didn't seem as if he was amused, and said: "If you want to learn how

    to drive, this is not the place to sit."

Subjects were presented with the following test sentence:

100. He gave me a/an ___ that didn't seem as if he was amused, and said: "If you want to learn

    how to drive, this is not the place to sit."

The most frequent word used to fill in the blank for this sentence was *look*, as opposed to stare.

Look is the noun found in the high-frequency exemplar LVC within the semantically similar

family including *give a stare*. For sentences involving test LVCs similar to low-frequency

exemplars, subjects filled in the blank with the noun of that low-frequency exemplar in just 0.3%

of the sentences (essentially one subject guessed the exact noun of the low-frequency exemplar

in one sentence). Thus, we can tentatively conclude that speakers have a high-frequency

exemplar in mind much more often than a low-frequency exemplar. Although this result is quite

different from that of Husain, Vasisth and Srinvasan (2014), who found that Hindi speakers were

able to correctly provide the verb for 86% of sentences involving LVCs, it should also be noted

that the task is quite different. Here, speakers are guessing the noun of the LVC based on

surrounding context and the light verb, which of course has very low cue validity. Furthermore,

these sentences were taken directly from the acceptability study, whereas the Hindi completion

study involved sentences intended to cue the listener for the sentence-final completion.

It's also noted above that it is possible for the sentential context to have cued more or less

strongly for a particular LVC combination or noun, regardless of the token frequency of the

combination. Essentially, acceptability may be due to the strength of the association between the

context and a particular LVC combination. To begin to examine this possibility, it was also noted

how often subjects were able to guess the noun of the actual test LVC (i.e. the noun found in the

original Gigaword sentence). For example, for sentence (99) above, it was noted how often

subjects guessed the noun from the actual Gigaword sentence, *stare*. For the sentences involving

test LVCs related to high-frequency exemplars, subjects filled in the blank with the noun from

the original Gigaword sentence (which includes the very low-frequency test LVC) in 13% of the

sentences. For sentences involving test LVCs related to low-frequency exemplars, subjects filled

in the blank with the noun from the original sentence in 10% of the sentences. Thus, it seems

that the extent to which the context cued for the test LVC was quite similar across the instances,

regardless of whether they were based on high or low-frequency exemplars. For the sentences

involving test LVCs based on low-frequency exemplars, it is notable that subjects were able to

guess the noun of that test LVC, found in the original sentence, much more often than they

guessed the low-frequency exemplar LVC. Although this followup merits further study, it

provides preliminary evidence that the high-frequency exemplar may indeed block the

acceptable usage of the very low-frequency LVC.

4.5.2 Semantic Similarity

The conclusions drawn regarding statistical preemption assume that the less acceptable, very

low-frequency LVCs had high-frequency counterparts that were recognized as preempting

structures. In order for this to occur, the two LVCs must be seen as semantically similar enough

to be competing variants, which are both equally suited to a particular context according to their

semantics, as described by Goldberg (2011). In this research, semantic similarity was firstly

based on FrameNet membership, and further ensured using LSA similarity. Nonetheless, it could

be argued that the less acceptable LVCs simply do not share the right type or level of semantic

similarity with their attested counterparts. For example, this work previously mentioned *make

speech, make proclamation, make announcement* vs. *?make yell.* It is quite easy to simply

dismiss *yell* as not having the "right" kind of semantic similarity to the attested examples. From

the perspective of categorization literature, it's plausible that the extension of LVC categorization

(and that of other semi-productive constructions) could be based on a variety of different factors

outside of simple semantic similarity of the noun.

Studies on radial categorization demonstrate the most comprehensive attempts for

pinpointing what is the "right" kind of similarity in extending category membership (Lakoff,

1987; Norvig & Lakoff, 1987; Michaelis & Ruppenhofer, 2001).   In past research by the present

author, manual sorting tasks (in which approximately 30 participants sorted *take* and *have* usages

into similar categories) have been used to develop a full radial categorization of both *take* and

have (unpublished, 2013). This research attempted to answer the question of why LVCs occur in

semantically similar families of phrases, yet their productivity seems idiosyncratically

constrained, and probed what "semantically similar" really means in this context. The preliminary results of the sorting task demonstrated that although the process of sorting into radial categories could be a valuable method for understanding the nature of semantic similarity involved in "families" of LVCs, the utility of the radial category approach may be limited because it's difficult to create a sorting task that is inclusive of all relevant usages (different instances for categorization would likely result in a different overall structure of categories), and radial categorization is necessarily subjective — different researchers surely arrive at different category structures and relations.

Furthermore, applying the notions of radial categorization fruitfully in the realm of NLP may be difficult. While there are many existing lexical resources that capture semantic relations between words, it would be difficult to capture the relations of interest in a radial categorization model since these are rather subjective, and the relations themselves are so varied and dependent on both context and culture. FrameNet does arguably come closest to capturing both semantic and pragmatic similarity, since it situates lexical items according to their shared membership in a real-world semantic domain. Thus, this research moves past the difficulties of establishing what kind of "similarity" may be important in extending constructions, while acknowledging that other factors may indeed affect the acceptability of particular LVC combinations.

Chapter 5
Broader Impacts & Future Work

This research brings a variety of disciplines to bear on English Light Verb Constructions: linguistics, psycholinguistics, computational linguistics and natural language processing, as well as cognitive science. Yet, there is more to be done to synthesize perspectives from each of these fields to come to a greater understanding of not only LVCs, but semi-productive, semi-compositional constructions generally. Impacts and future work possibilities relating to each research question are discussed in turn in the sections to follow.

5.1 Research Question 1: Definition and Delimitation of LVCs

The study of English LVCs, and LVCs cross-linguistically, is hampered by a lack of consensus in the linguistic community on what precisely defines such constructions, how they can be delimited from other constructions in the language, as well as how they should be labeled (light/ support verbs, composite predicates, etc.). Language-dependent criteria are needed, as each language has unique resources for expressing eventualities and participants, and these resources also interact uniquely. While clear syntactic criteria are extremely desirable in defining and delimiting LVCs, it seems that such criteria are simply not available in some languages, including English, in which LVCs and compositional *verb + noun* usages are quite often indistinct. Thus, after existing definitions and definitional criteria (both syntactic and semantic) were discussed, Chapter 2 delved into original work exploring primarily semantic criteria for defining LVCs and distinguishing them from heavy usages, idioms, and other support verb types.

177

This discussion culminated in a definition of LVCs that is partially syntactic, but largely semantic. The basic criteria for an LVC in English are:

1. monoclausal structure

2. verb of semantically general meaning

3. noun complement that is abstract, denoting an event or state

Although the definition is somewhat underspecified, the heuristics and array of positive and negative examples provided to PropBank annotators in the guidelines have proven to be effective, as demonstrated in the high agreement rates between annotators on likely LVC annotation tasks.

Annotations that are free of unsystematic errors (which would be reflected in low agreement rates) in turn make for good training data. The quality of PropBank LVC annotations as training data has been demonstrated in the development of the state-of-the-art system for LVC identification, which achieves an F-Score of 80% when tested on the PropBank corpus. In a comparison to the previous benchmarks in LVC detection, we see that the PropBank corpus of LVCs is extremely comprehensive because it includes a wide variety of potential light verbs and LVC noun complements, where much other work restricts these to a certain subset of LVC combinations. While such restrictions do simplify the problem, they can also eliminate clearly related phenomena, such as LVCs involving eventive or stative nouns with no clear verbal counterparts (e.g. *take a trip*). The completeness of the PropBank annotation schema reflects the depth of theoretical research that underlies it, and brings a consistent and comprehensive delimitation of LVCs where previously definitional inconsistencies precluded this.

However, a comparison to other LVC annotation resources, specifically that of the TectoGrammatical Treebank (TTree), also shows that PropBank annotations may have gaps or inconsistencies in how the larger category of support verbs are annotated. The definition and guidelines outlined in PropBank allow us to distinguish various types of support verbs, but we should also recognize their commonalities: all support verbs syntactically support the realization of a subject argument shared by the verb and relational noun. Therefore, in future work, we should explore treating Causal and Aspectual support verbs in a manner more similar to LVCs, so that the full range of arguments of a predicating noun complement can be consistently expressed and annotated. Additionally, the comparisons to other annotation resources in this research revealed that we should certainly loosen the definition of LVCs to include what are termed here 'semi-LVCs,' because the fine-grained distinctions between LVCs and semi-LVCs currently being made are leading to problematic inconsistencies in annotations: an LVC involving one light verb (e.g. *give a speech*) is not considered truly light, but an LVC involving another light verb (e.g. *make a speech*) is considered light. For the purposes of NLP, it is chiefly important to consistently recognize both of these as *speech* events so that the appropriate inferences are made across the various possible syntactic realizations of the same event.

Furthermore, although the state of the art in detecting LVCs is advancing, another goal essential to Natural Language Understanding is the correct interpretation of LVCs. The PropBank annotation schema provides some shallow semantics of LVCs in the form of semantic role assignments — it is clear who is doing what to whom, as well as how, where, when and why if this is specified. A sister project to PropBank, VerbNet (Kipper et al., 2008), provides a deeper level of semantic analysis in the form of semantic predicates. SemLink (Palmer, 2009; Loper et

al., 2007), which provides mappings between PropBank and VerbNet, allows for VerbNet

predicates to be associated with annotated PropBank instances. However, in the case of LVCs,

this mapping is not so simple because the relation marked in the PropBank instance is a noun,

and nouns are not part of VerbNet. Nonetheless, many of these nouns have etymologically

related or semantically similar lexical verb counterparts that do have membership in VerbNet.

Current work to unify PropBank frame files across varying morphosyntactic realizations of a

given concept (e.g. one unified frame file for *fear*-verb, *fear*-noun, and *afraid*-adjective) makes

the relationship between eventive and stative nouns and their verbal counterparts explicit (Bonial

et al., 2014a), thereby also providing an efficient mapping between PropBank LVCs annotated

with noun rolesets, and the corresponding VerbNet information for that instance.

However, interpretation using VerbNet would rely upon the notion that the semantics of

the LVC can be adequately captured by essentially replacing the LVC with a verb related to (or

semantically very similar to) its head noun predicate. This is also the approach of the Abstract

Meaning Representation (AMR) project, which attempts to move beyond language-specific

syntactic facts to represent the core facets of meaning (Banarescu et al., 2013). Given this goal,

it seems appropriate that the AMR project represents LVCs by stripping away the light verb

(which can be thought of as a syntactic idiosyncrasy of English) and replacing the entire

construction with a lexical verb that captures the semantics of the event or state denoted by the

noun. Although this approach may be adequate for the interpretation of LVCs in English, it

seems somewhat unlikely that it would be an appropriate cross-linguistic treatment of LVCs,

since there is a cross-linguistic tendency for LVCs to cover the semantic space that is not covered

by lexical verbs. Certainly further exploration into using NLP resources for the correct

interpretation of LVCs is needed, but we have taken the preliminary steps for defining and

detecting LVCs, allowing for a focus on interpretation in the future.

5.2 Research Question 2: LVC Function

English LVCs (and those of many Romance languages) are unique in that they do tend to have

lexical verb counterparts that are semantically quite similar. For example:

102. She **appeared** with me on VH1 "Celebrity Rehab."
103. This afternoon, Bahrains King Hamad **made** a rare **appearance** on television.

This fact runs contrary to many assumptions in linguistics that two competing forms are rarely

maintained in the language, unless they serve distinct purposes. Therefore, why do LVCs exist

alongside counterpart lexical verbs, especially given that lexical verbs are arguably the more

efficient variant form, and the semantics of the two forms are quite similar? In the survey of

diachronic and synchronic viewpoints on the topic, we saw two important factors: aspectual

differences between LVCs and counterpart verbs, and differences in the ease of modification of

LVCS and counterpart verbs.

　　　　To examine the validity of these views in today's English, a pilot corpus study was firstly

undertaken, leading to the development of the Modification Hypothesis:

> **Modification Hypothesis:** Speakers elect to use an LVC of Type 3 (with predicating
> nouns that are not related to a verb) out of necessity — simply because a semantically
> similar counterpart verb doesn't exist — but speakers use LVCs of Types (1) and (2)
> (with predicating nouns that are derivationally or etymologically related to a verb), as

opposed to their verbal counterparts, because they want to convey additional descriptions of the event, facilitated by nominal expression.

To test this hypothesis, the full corpus study described herein was undertaken, using computational techniques to extract and compare the PropBank annotations of nearly 2,000 LVCs, 10,000 counterpart verbs, and about 18,000 eventive and stative nouns outside of LVCs. Bear in mind that this theoretical study would not have been possible without the years of annotation effort that contributed to the development of PropBank LVC annotations. The corpus study firstly examined aspectual differences between the variant forms, and secondly the differences in the quantities and qualities of modification. While the data did not demonstrate that there is any consistent relationship between the aspect of the LVC and the aspect of its counterpart lexical verb, the data did show that modification generally seems to be the motivating factor for speakers to elect to use an LVC over a counterpart lexical verb, supporting the Modification Hypothesis. Specifically, LVCs are much more often modified than counterpart lexical verbs: there are, on average, 1.15 modifiers per LVC and 0.6 modifiers per counterpart lexical verb. Notably, this modification can include certain types of determiners (or lack thereof) that allow speakers to modulate aspectual meaning of the construction. Thus, we see that both aspect and modification more generally are both important factors.

While this is a distributional account of the function of LVCs, it does not necessarily indicate direct causation, in that we still cannot be certain that a speaker's need to express certain descriptive elements would motivate the use of an LVC over a counterpart lexical verb. Thus, an experimental paradigm examining whether the need to supply certain modificational elements prompts speakers to use an LVC over a counterpart lexical verb would be useful. However,

because LVCs often convey somewhat idiomatic, and certainly more abstract types of events, such a paradigm seems very difficult to achieve and we must, for now, settle for the distributional evidence.

It is also important to note that while ease of modification seems to be the most general motivating force for maintaining LVCs in English alongside counterpart lexical verbs, individual LVCs can also take on nuances of meaning and pragmatics that make their usage distinct from counterpart lexical verbs. While the question of what such individual differences are is an important one, this dissertation provided evidence from existing research (Wierzbicka, 1982; Lakoff, 1987; Norvig & Lakoff, 1987) that pinpointing such differences on the scale that would be required for NLP may be intractable. Nonetheless, research on a larger scale exploring the nuances of meaning contributed by the LVC could bring to light patterns of how individual light verbs tend to modulate the meaning of the event or state designated by the noun. The preliminary steps towards such research have already begun — using the growing number and variety of PropBank LVC annotations, I have started to investigate and tally the following features of light verbs: 1) which light verbs tend to co-occur with the same eventive/stative nouns, 2) what is the semantic role of the syntactic (subject) argument for the clusters of light verbs detected in (1). With these two relatively simple pieces of information, we should be able to begin to see patterns of light verb usages.

For example, the twenty most frequent LVCs involving *make* were extracted, and the nouns from these LVCs were entered into the Word Sketch function of Sketch Engine , which then outputs the most frequent verb collocates of that noun in the large English TenTen corpus (Jakubíček, 2013). When these verb collocates are examined across all of the nouns from the

183

twenty *make* LVCs, we see that the potential light verb *receive* occurs more frequently than any

other verb with the same noun collocates as *make*. However, an examination of usages shows

that *make* and *receive* link distinct subject arguments. For example, *make payment* links an

Agent subject, while the semi-LVC *receive payment* links a Recipient subject. While certain

pairs of LVCs are well-known in linguistics (e.g. *take a bath* vs. *give a bath*), the PropBank LVC

annotations may allow for the first large-scale exploration of the tendencies of light verbs. Such

a project would also assist in gaining a better understanding of what type of system will be

needed for interpretation of LVCs: to what extent does correct interpretation rely on

understanding the LVC as a combination of a verb and noun, as opposed to understanding the

LVC as the event denoted by the noun alone?

5.3 Research Question 3: LVC Productivity

LVC productivity is perhaps the most linguistically fascinating and practically important

question addressed in this research. To understand how semi-productive constructions like LVCs

are extended and how that productivity is constrained provides evidence for how grammar

functions generally, as an emergent structure that arises out of a speaker's experience with each

token of linguistic experience. Emergent Grammar predicts that novel constructions are

extended by analogy to high-frequency, existing constructions, and this has been demonstrated in

very similar research on Spanish *becoming* constructions (Bybee & Eddington, 2006). What this

research has not generally addressed is what is meant by "high-frequency." We know that the

existing structure must be well established in the speaker's grammar for it to serve as the basis of

extension, but the Exemplar Model also predicts that relatively few experiences with a given

exemplar may be enough for a speaker to be familiar with it, since a wide variety of linguistic details are stored with each exemplar.

In this research, this view is tested in the domain of LVCs using carefully crafted surveys of LVC acceptability administered on a large-scale using Amazon's Mechanical Turk. The hypothesis of Bybee & Eddington (2006) is adapted to this work:

> **Frequency Hypothesis:** Speakers will find novel or very low-frequency LVCs acceptable if they are semantically similar to an attested, highly frequent LVC.

Subjects in the study were asked to judge the acceptability or naturalness of very low-frequency LVCs (in original Gigaword context) that were semantically similar (according to FrameNet membership of the noun) to either a high-frequency LVC exemplar or a low-frequency LVC exemplar. Much existing work in Emergent Grammar and Construction Grammar would predict that speakers will find the test LVCs that are semantically similar to a high-frequency exemplar to be more acceptable than those that are semantically similar to a low-frequency exemplar.

In fact, the frequency of the exemplar LVC does significantly correlate with the acceptability of the very low-frequency test LVC, but in the opposite direction: very low-frequency LVCs (20-50 instances in Gigaword) that are semantically similar to a low-frequency LVC (100-200 instances) are 8.56 points (on a scale from 0-100, +/- 3.50 standard errors) more acceptable than very low-frequency LVCs that are semantically similar to a high-frequency exemplar (2,000-3,000 instances) ($X^2(1)$=5.64, p=0.01751). This finding does not necessarily undermine Bybee's theories as to how constructions are extended (2006, 2010; Bybee & Eddington, 2006), but instead elaborates upon this research in several important ways. Firstly, it demonstrates that relatively low-frequency constructions can indeed serve as the basis of

analogical extensions. Secondly, it shows that there is a certain point where the higher frequency of a semantically similar form may block another variant, rather than encouraging analogical extension. In other words, this research provides evidence for the process of statistical preemption: speakers implicitly infer from consistently hearing a formulation, B, in a context where one might have heard a semantically related alternative formulation, A , that B is the appropriate formulation and A is not appropriate.

This research then explores other linguistic work in the area of statistical preemption, and attempts to bridge the gap between this linguistic work and potentially related work in both psycholinguistics and cognitive science. Recent work exploring the "semantic P600" (e.g. Kim & Sikos, 2011) suggests that subjects complete argument structure patterns according to frequency-based expectations, and observed completions that conflict with the speaker's expectations may seem syntactically anomalous. Additionally, it has been shown that there are processing difficulties associated with "shattered expectations" (Husain, Vasisth and Srinivasan, 2014). These findings may be related to why subjects find the marked, very low-frequency LVC variants of high-frequency exemplars "odd." Specifically, the results of testing the Frequency Hypothesis, as well as the preliminary evidence from the followup study, show that speakers have an expectation for the high-frequency variant, and when the very low-frequency variant is encountered in the context that seems to be equally appropriate for the high-frequency variant, the very low-frequency LVC seems odd. Examining statistical preemption in online processing (perhaps using an ERP study) is a fruitful area for future work. While the online processing of LVCs, or at least constructions involving vague action verbs, has been explored in some detail

(e.g. Piñango et al., 2006; Wittenberg et al., submitted), this type of research has not yet begun to explore frequency effects in relation to LVC processing.

In other future work, the intermediate levels of frequency should be further explored, in an effort to determine if there seems to be any identifiable threshold of frequency for where speakers allow for analogical extensions, and a separate threshold of frequency where preemption begins to occur instead of extension. Such research may also help to examine the potential impacts of this finding on NLP. This research explicates, to some extent, why the "families" of LVCs seem to be so idiosyncratic — there is an interplay of extension and preemption — but it does not provide the type of information that could be encoded in an automatic system to weight the probability of a given combination arising as an acceptable LVC. If such frequency bands of extension and preemption could be determined, then perhaps this information could be leveraged to determine what families of semantically similar LVCs might be "open" to new members, and which already contain many high-frequency LVCs that will potentially preempt new members.

5.4 Concluding Thoughts

As a whole, this research has demonstrated the promising ways in which the various disciplines of linguistics, computer science and cognitive science can be brought together to identify a linguistic phenomenon, understand its function in the language, and begin to understand how humans can, and perhaps computers should, process the phenomenon. The focus on English LVCs has been particularly challenging because LVCs are both semi-compositional and semi-

productive, making their treatment in the mental lexicon, as well as computational lexicons, problematic. Nonetheless, what is most challenging is also most rewarding, and this work has demonstrated the importance of understanding semi-productive, semi-compositional constructions. For if we can use research in psycholinguistics and cognitive science to better understand how humans learn the idiosyncratic constraints of such constructions, including how statistical preemption plays into speakers' knowledge of these constraints, then this understanding would assist in training computers to learn such constraints. The successful computational modeling of such constraints would, in turn, lay the groundwork for a superior understanding of grammar as a whole, and the effects of frequency on grammar.

References

Abeillé, Anne, Danièle Godard, and Ivan A. Sag. 1998. Two Kinds of Composition in French Complex Predicates. In *Complex Predicates in Nonderivational Syntax*, ed. Erhard Hinrichs, Andreas Kathol, and Tsuneko Nakazawa. San Diego: Academic Press. *Syntax and Semantics* Volume 30.

Adger, D. 2003. *Core syntax: a Minimalist approach.* Oxford: Oxford University Press.

Akimoto, Minojo. 1999. Collocations and Idioms in Late Modern English. In Brinton, L.J. and Akimoto, M. (Eds) *Collocational and Idiomatic Aspects of Composite Predicates in the History of English.* Philadelphia: John Benjamins.

Akimoto, Minojo. and Brinton, Laurel J. 1999. Origin of the composite predicate in Old English. In Brinton, L.J. and Akimoto, Minojo. (Eds) *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*. Philadelphia: John Benjamins.

Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. *Proceedings of the 17th International Conference on Computational Linguistics (COLING/ACL-98)*, pp. 86– 90, Montreal.

Banarescu, L., Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer and Nathan Schneider. Abstract Meaning Representation (AMR) 1.0 Specification. http://www.isi.edu/~ulf/amr/help/amr-guidelines.pdf.

Barrett, L. and A.R. Davis. 2003. Diagnostics for determining compatibility in English support-verb-nominalization pairs. *Computational Linguistics and Intelligent Text Processing,* 85-90.

Barsalou, Lawrence W. 1985. Ideals, Central Tendency, and Frequency of Instantiation as Determinants of Graded Structure in Categories. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 11(4), 629-649.

Bates D, M. Maechler, B. Bolker and S. Walker. 2014. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7, http://CRAN.R-project.org/package=lme4.

Bermel, Neil and Luděk, Knittl. 2012a. Corpus Frequency and Acceptability: A study of morphosyntactic variants in Czech. In Gries, Stefan (Ed) *Corpus Linguistics and Linguistic Theory,* Volume 8, Issue 2, pp. 241-275.

Bermel, Neil and Luděk, Knittl. 2012b. Morphosyntactic variation and syntactic constructions in Czech nominal declension: corpus frequency and native-speaker judgments. *Russian Linguistics,* Volume 36, Issue 1, pp. 91-119.

Bonial, Claire, Susan Windisch Brown, William Corvey, Martha Palmer, Volha Petukhova, and Harry Bunt. 2011a. An Exploratory Comparison of Thematic Roles in VerbNet and LIRICS. In *Proceedings of the Sixth Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation (ISA-6)*, 39-44, Jan, 2011.

Bonial, Claire, William Corvey, Martha Palmer, Volha Petukhova, and Harry Bunt. 2011b. A Hierarchical Unification of LIRICS and VerbNet Semantic Roles. In *Proceedings of the ICSC Workshop on Semantic Annotation for Computational Linguistic Resources (SACL-ICSC 2011)*, Sep, 2011.

Bonial, Claire, Susan Windisch Brown, Jena D. Hwang, Christopher Parisien, Martha Palmer and Suzanne Stevenson. 2011c. Incorporating Coercive Constructions into a Verb Lexicon. In *Proceedings of the ACL 2011 Workshop on Relational Models of Semantics (RELMS'11)*, Portland, Oregon. June, 2011.

Bonial, Claire, Julia Bonn, Kathryn Conger, Jena D. Hwang & Martha Palmer. 2013. *English PropBank Annotation Guidelines.* http://verbs.colorado.edu/propbank/EPB-Annotation-Guidelines.pdf.

Bonial, Claire, Julia Bonn, Kathryn Conger, Jena D. Hwang & Martha Palmer. PropBank: Semantics of New Predicate Types. 2014a. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14),* pp. 3013-3019.

Bonial, Claire, Meredith Green, Jenette Preciado & Martha Palmer. 2014b. An Approach to Take Multi-Word Expressions. In *Proceedings of the EACL 10th Workshop on Multiword Expressions (MWE 2014)*.

Bowern, Claire. 2002. How light are North Australian 'Light Verbs'? Talk given at the *Light Verb Workshop*, Harvard University, November 2002.

Boyd, Jeremy K., Farrell Ackerman and Marta Kutas. 2012. Adult learners use both entrenchment and preemption to infer grammatical constraints. In *Proceedings of the 2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)*.

Briem, Daniela, Britta Ballel, Brigitte Rockstroh, Miriam Butt, Sabine Schulte im Walde, Ramin Assadollahi. 2009. Distinct processing of function verb categories in the human brain. *Brain Research* vol. 1249. pp 173-180.

Brinton, Laurel J. & Arnovick, Leslie K. 2011. *The English Language: A Linguistic History.* Canada: Oxford University Press.

Brinton, Laurel J. and Akimoto, Minojo, editors. 1999. *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*. Philadelphia: John Benjamins Publishing Company.

Broadwell, Aaron G. 2000. Choctaw directionals and the syntax of complex predication. *Argument Realization.* Eds. Miriam Butt and Tracy King. Stanford: CSLI Publications.

Brooks, L. 1978. Nonanalytic concept formation and memory for instances. In E. Rosch and B.B. Lloyds (eds.) *Cognition and Categorization.* Hillsdale, NJ: Erlbaum.

Brown, Susan Windisch, Dmitriy Dligach, and Martha Palmer. VerbNet Class Assignment as a WSD Task. In *IWSC 2011: Proceedings of the 9th International Conference on Computational Semantics,* January 12 - 14, 2011, Oxford, UK.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Dissertations in Linguistics. Stanford, CA: CSLI Publications.

Butt, Miriam. 2003. The Light Verb Jungle. In G. Aygen, C. Bowern & C. Quinn (Eds.) *Papers from the GSAS/ Dudley House Workshop on Light Verbs*. Cambridge, Harvard Working Papers in Linguistics, p. 1-50.

Butt, Miriam, and Wilhelm Geuder. 2001. On the (Semi)Lexical Status of Light Verbs. In *Semi-Lexical Categories: On the content of function words and the function of content words*, ed. Norbert Corver and Henk van Riemsdijk. 323-370. Berlin: Mouton De Gruyter.

Bybee, Joan. 2006. From usage to grammar: the mind's response to repetition. *Language* 82(4): 711-733.

Bybee, Joan. 2010. Language, Usage and Cognition. Cambridge: Cambridge University Press.

Bybee, Joan, and D. Eddington. 2006. A usage-based approach to Spanish verbs of 'becoming'. *Language*, 82(2), 323–355.

Bybee, Joan. and Paul Hopper. 2001. Introduction. In J. Bybee and P. Hopper (Eds) Frequency and the Emergence of Linguistic Structure. Amsterdam: John Benjamins.

Bybee, Joan, and J. McClelland. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. The Linguistic Review 22: 381-410.

Cattel, Ray. 1984. *Syntax and Semantics: Composite Predicates in English.* Orlando: Academic Press, Inc.

Chen, Wei-Te, Claire Bonial & Martha Palmer. English Light Verb Construction Identification Using Lexical Knowledge. Submitted to the *Twenty-Ninth Conference on Artificial Intelligence (AAAI-15)*.

Choi, Jinho D., Claire Bonial & Martha Palmer. 2009. *Jubilee: Propbank Instance Editor Guideline (Version 2.1)*, Technical report 02-09: Institute of Cognitive Science, University of Colorado at Boulder, Boulder, CO.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass: MIT Press.

Chomsky, Noam. 1980. *Rules and representations*. New York: Columbia University Press.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordecht: Foris.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Mass.: MIT Press.

Church, Kenneth Ward and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. Computational Linguistics, 16 (1): 22–29.

Church, Kenneth Ward, Patrick Hanks, and D. Hindle. 1991. *Using Statistics in Lexical Analysis*. p. 115-164. Lawrence Erlbaum.

Cinková et al. 2004. *Annotation of English on the tectogrammatical level*. Technical Report TR-2006- 35, ´UFAL/CKL, Prague, Czech Republic.

Croft, William and D.A. Cruse. 2004. From Idioms to Construction Grammar. Chapter 9, *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Crystal, David. 2003. *A Dictionary of Linguistics & Phonetics- 5th Edition*. Padstow, Cornwall: Blackwell Publishing.

Cutting, J. Cooper and Kathryn Bock. 1997. That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition,* 25(1): 57-71.

Davies, M. 2008-. The Corpus of Contemporary American English (COCA): 425 million words, 1990-present. Available online at http://www.americancorpus.org.

Davis, A.R. and L. Barrett. 2001. Support Verbs, Nominalizations and Participant Roles. In *Proceedings of the First International Workshop on Generative Approaches to the Lexicon*. University of Geneva, Geneva.

Divjak, Dagmar. 2008. On (in)frequency and (un)acceptability. In Barbara Lewandowska-Tomaszczyk (ed.), *Corpus linguistics, computer tools and applications – State of the art*, pp. 213–233. Frankfurt: Peter Lang.

Dligach, Dmitriy and Martha Palmer. 2011. Good Seed Makes a Good Crop: Accelerating Active Learning Using Language Modeling. In *ACL '11: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.

Dras, M. and M. Johnson. 1996. Death and lightness: Using a demographic model to find support verbs. In *5th International Conference on the Cognitive Science of Natural Language Processing*.

Duffield, Cecily Jill, Jena D. Hwang; Susan Windisch Brown; Dmitriy Dligach; Sarah E. Vieweg; Jenny Davis; Martha Palmer. 2007. Criteria for the Manual Grouping of Verb Senses. *Proceedings of the Linguistic Annotation Workshop held in conjunction with ACL-2007*.

Fazly, Afsaneh. 2007. *Automatic Acquisition of Lexical Knowledge about Multiword Predicates*. PhD Thesis, Department of Computer Science, University of Toronto.

Fazly, Afsaneh, Ryan North, and Suzanne Stevenson. 2005. Automatically Distinguishing Literal and Figurative Usages of Highly Polysemous Verbs. In *Proceedings of the ACL 2005 Workshop on Deep Lexical Acquisition*, 38-47. Ann Arbor, Michigan, USA, June, 2005.

Fazly, Afsaneh and Suzanne Stevenson. 2006. Automatically Constructing a Lexicon of Verb Phrase Idiomatic Combinations. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, pp. 337-344.

Fazly, Afsaneh and Suzanne Stevenson. 2007. Distinguishing subtypes of multiword expressions using linguistically-motivated statistical measures. In *Proceedings of the ACL'07 Workshop on A Broader Perspective on Multiword Expressions*, pp. 9-16.

Fellbaum, C. (Ed.) 1998. *WordNet: An Electronic Lexical Database.* MIT Press.

Fillmore, Charles J., and Paul Kay. 1999. Grammatical constructions and linguistic generalizations: the What's X doing Y? construction. *Language* 75(1): 1–33.

Fillmore, Charles J., Christopher R. Johnson, and Miriam R.L. Petruck. 2002. Background to FrameNet. *International Journal of Lexicography*, 16(3):235-250.

Frajzyngier, Zygmunt and Erin Shay. 2003. *Explaining language structure through systems interaction.* Amsterdam and Philadelphia: Benjamins.

Fraser, Bruce. 1970. Idioms Within a Transformational Grammar. *Foundations of Language,* 6(1): 22-42.

Givón, Talmy. 1979. On understanding grammar. New York: Academic Press.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.

Goldberg, Adele. 1996. Construction Grammar. In Keith Brown and Jim Miller (eds.), *Concise Encyclopedia of Syntactic Theories*. New York: Elsevier, 68–71.

Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language.* New York: Oxford University Press.

Goldberg, Adele. 2011. Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics* 22 1: 131-154.

Graff, David and Christopher Cieri. 2003. *English Gigaword.* Linguistic Data Consortium, Philadelphia.

Grefensette, G. and S. Teufel. 1995. A corpus-based method for automatic identification of support verbs for nominalisations. In *Proceedings of EACL,* Dublin, Ireland, pp. 98-103.

Grice, H.P. 1975. Logic and Conversation. In P. Cole and J. Morgan (Eds.) *Syntax and Semantics*, vol.3. Academic Press. Reprinted as ch.2 of Grice 1989, 22–40.

Grimshaw, J., and A. Mester. 1988. Light verbs and θ-marking. *Linguistic Inquiry,* 19(2):205–232.

Gurevich, Olga, Matt Johnson, and Adele E. Goldberg. 2010. Incidental verbatim memory for language. *Language and Cognition*. Cambridge University Press.

Hagoort, P., C. Brown, and J. Groothusen. 1993. The syntactic positive shift as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439–483.

Hartshorne, Joshua, Claire Bonial and Martha Palmer. 2013. The VerbCorner Project: Toward an Empirically-Based Semantic Decomposition of Verbs. In *Proceedings of EMNLP-13,* pp. 1438-1442.

Hiltunen, Risto. 1999. Verbal Phrases and Phrasal Verbs in Early Modern English. In Brinton, L.J. and Akimoto, M. (Eds) *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*. Philadelphia: John Benjamins.

Hintzman, D.L. 1986. "Schema abstraction" in a multiple-trace model. *Psychological Review,* 93, 411-428.

Hopper, Paul J. 1987. Emergent grammar. In J. Aske, N. Beery, L. Michaelis, and F. Hana (Eds) *Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society*.

Hopper, Paul J. 1991. On Some Principles of Grammaticalization. *Approaches to Grammaticalization,* Vol. 1: *Focus on Theoretical and Methodological Issues* (=*Typological Studies in Language,* 19:1) ed. by Elizabeth Closs Traugott and Bernd Heine, 17-35. Amsterdam & Philadephia: John Benjamins.

Hopper, Paul J. 1998. Emergent Grammar. In *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*, Michael Tomasello (ed.), 155-75. Mahwah, NJ/London: Lawrence Erlbaum Associates.

Hopper, Paul J. and Elizabeth C. Traugott. 1993. *Grammaticalization (=Cambridge Textbooks in Linguistics.)* Cambridge: Cambridge University Press.

Horn, Laurence. 1985. Towards a new taxonomy for pragmatic inference: Q-based and R-based implicature. Meaning, form, and use in context, ed. by Deborah Schiffrin, 11–42. Washington, DC: Georgetown University Press.

Huang, Charles T. 1992. Complex Predicates in Control. In *Control and Grammar,* ed. Richard Larson, Sabine Iatridou, and Utpal Lahiri. 109-147. Kluwer Academic Publishers.

Husain, Samar, Shravan Vasisth, Narayan Srinivasan. 2014. Strong Expectations Cancel Locality Effects: Evidence from Hindi. *PLOS one,* Volume 9, Issue 7.

Hwang, Jena D, Archna Bhatia, Claire Bonial, Aous Mansouri, Ashwini Vaidya, Nianwen Xue, and Martha Palmer. 2010a. PropBank Annotation of Multilingual Light Verb Constructions. In Proceedings of the Linguistic Annotation Workshop held in conjunction with ACL-2010. Uppsala, Sweden, July 15-16, 2010.

Hwang, Jena D., Rodney D. Nielsen and Martha Palmer. Toward a Domain Independent Semantics: Enhancing Semantic Representation with Construction Grammar. In Proceedings of Extracting and Using Constructions in Computational Linguistics Workshop, held in conjunction with NAACL-HLT 2010, June 6, 2010, Los Angeles, CA.

Israel, Michael. 1996. The way constructions grow. In A. E. Goldberg (ed.), *Conceptual structure, discourse, and language*, 217–30. Stanford, CA: CSLI.

Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý, Vít Suchomel. 2013. The TenTen Corpus Family. In *Proceedings of the 7th International Corpus Linguistics Conference*, Lancaster, July 2013.

Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51: 639-671.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

Jakobson, Roman. 1990. Some questions of meaning. In L. R. Waugh (ed.), *On language: Roman Jakobson*, 315–23. Cambridge, MA: Harvard University Press.

Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles.* Part VI: *Morphology*. With assistance of P. Christopherson, N. Haislund, & K. Schibsbye. London: Goerge Allen & Unwin. Copenhagen: Ejnar Munksgaard.

Johnson, Keith. 1997. Speech perception without speaker normalization. In K. Johnson and J. W. Mullennix (eds.), *Talker variability in speech processing*, 145–65. San Diego, CA: Academic Press.

Kastovsky, Dieter. 1985. Deverbal nouns in Old and Modern English: from stem-formation to word-formation. In J. Fisiak (ed.) *Historical Semantics, Historical Word-Formation*. 221-261. Berlin: Mouton Publishers.

Katz, Jerrold J. and Paul Postal. 1963. Semantic interpretation of idioms and sentences containing them. Quarterly Progress Report of the MIT Research Laboratory of Electronics 70. 275-82.

Katz, G. and E. Giesbrecht. 2006. Automatic identification of non-compositional multi-word expressions using latent semantic analysis. In *Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pp 12–19.

Kay, Paul and Ivan Sag. 2012. *A Lexical Theory of Phrasal Idioms*. Unpublished manuscript, Stanford University. http://www1.icsi.berkeley.edu/~kay/idioms-submitted.pdf.

Kearns, Kate. 2002. *Light verbs in English*. http://www.ling.canterbury.ac.nz/kate/lightverbs.pdf.

Kilgarriff, Adam, Pavel Rychly, Pavel Smrz, and David Tugwell. 2004. The Sketch Engine. *Proceedings of EURALEX*. Lorient, France.

Kim, A., and L. Osterhout. 2005. The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language,* 52(2), pp. 205-225.

Kim, A. and L. Sikos. 2011. Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain & Language*, 118, 15-22.

Kipper, Karin, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation Journal*, 42: pp. 21– 40.

Kutas, M., & S.A. Hillyard. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, pp. 203–205.

Kyto, Merja. 1999. Collocational and Idiomatic Aspects of Verbs in Early Modern English: A corpus-based study of MAKE, HAVE, GIVE, TAKE and DO. In Brinton, L.J. and Akimoto, M. (Eds) *Collocational and Idiomatic Aspects of Composite Predicates in the History of English.* Philadelphia: John Benjamins.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things.* Chicago: University of Chicago Press.

Landauer, Thomas, Peter W. Foltz, and Darrell Laham. 1998. Introduction to Latent Semantic Analysis. *Discourse Processes* 25 (2–3): 259–284.

Langacker, Ronald. 2000. A dynamic usage-based model. In Michael Barlow & Suzanne Kemmer (eds.), *Usage-based models of language*, 1–63. Stanford, CA: CSLI Publications.

Lee, Y. K., and H.T. Ng. 2002. An empirical evaluation of knowledge sources and learning algorithms for word sense disambiguation. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10, EMNLP '02*, 41–48.

Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Lin, D. 1999. Automatic identification of non-compositional phrases. In *Proceedings of ACL-99,* pp. 317-324.

Live, Anna H. 1973. The *take-have* Phrasal in English. *Linguistics,* 9: 31-50.

Loper, Edward, Szu-ting Yi and Martha Palmer. Combining Lexical Resources: Mapping Between PropBank and VerbNet. *Proceedings of the 7th International Workshop on Computational Semantics*. Tilburg, the Netherlands. 2007.

MacWhinney, B. 2001. Emergentist approaches to language. In J. Bybee and P. Hopper (Eds). Frequency and the Emergence of Linguistic Structure, pp. 449-470. Amsterdam: John Benjamins.

Malt, Barbara C. 1989. An On-Line Investigation of Prototype and Exemplar Strategies in Classification. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 15(4), 539-555.

Manning C.M., and H. Schütze. *Foundations of statistical natural language processing*. Cambridge, MA: The MIT Press; 1999.

Matsumoto, Meiko. 1999. Composite Predicates in Middle English. In Brinton, L.J. and Akimoto, M. (Eds) *Collocational and Idiomatic Aspects of Composite Predicates in the History of English.* Philadelphia: John Benjamins.

Marcus, M., B. Santorini and M. Marcinkiewicz. 1993. Building a Large Annotated Corpus of English: the Penn Treebank. *Computational Linguistics,* 19 (2): 313-330.

Màrquez, L., X. Carreras, K. Litkowski, and S. Stevenson. 2008. Semantic role labeling: An introduction to the special issue. *Computational Linguistics*, 34(2): 145-159.

Mathesius, Vilem. 1975. *A functional Analysis of Present-Day English on a General Linguistic Basis* (=*Janua Linguarum, Series Practicum,* 208). The Hague: Mouton.

McCarthy, Diana, Bill Keller, and John Carroll. 2003. Detecting a Continuum of Compositionality in Phrasal Verbs. In *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, Sapporo, Japan.

Medin, D.L. and M.M. Schaffer. 1978. Context theory of classification learning. *Psychological Review,* 85, 207-238.

Meyers, Adam, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. 2004. The NomBank Project: an Interim Report. *Frontiers in Corpus Annotation, Workshop in conjunction with HLT/NAACL.*

Michaelis, Laura A. and Josef Ruppenhofer. 2001. *Beyond Alternations: A Constructional Model of the German Applicative Pattern.* Stanford: CSLI Publications.

Michaelis, Laura A. 2004. Type Shifting in Construction Grammar: An Integrated Approach to Aspectual Coercion. *Cognitive Linguistics* 15: 1-67.

Kay, Paul and Laura A. Michaelis. 2012. Constructional Meaning and Compositionality. In C. Maienborn, K. von Heusinger and P. Portner (eds.), *Semantics: An International Handbook of Natural Language Meanin*g. Vol. 3. Berlin: de Gruyter. 2271-2296.

Miller, George A. 1995. WordNet: A Lexical Database for English. *Communications of the ACM* Vol. 38, No. 11: 39-41.

Moens, M. and M. Steedman. 1988. Temporal Ontology and Temporal Reference. *Computational Linguistics*, 14:15-38.

Mohanan, Tara. 1994. *Argument Structure in Hindi.* Stanford, California: CSLI Publications.

Nickel, Gerhard. 1978. Complex Verbal Structures in English. *Studies in Descriptive English Grammar,* 63-83.

North, Ryan. 2005. *Computational Measures of the Acceptability of Light Verb Constructions.* MA Thesis, University of Toronto.

Norvig, Peter and George Lakoff. 1987. Taking: A Study in Lexical Network Theory. In *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society.* 195-206.

Novick, J., Kim, A.E., & Trueswell, J. 2003. Studying the grammatical aspects of word recognition: Lexical priming, parsing and syntactic ambiguity resolution. *Journal of Psycholinguistic Research* 32(1), 57-75.

Nunberg, G., Sag, I.A., & Wasow, T. 1994. Idioms. *Language*, 70: p. 491-538.

Palmer, Martha. 2009. Semlink: Linking PropBank, VerbNet and FrameNet. *Proceedings of the Generative Lexicon Conference*.

Palmer, Martha, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71– 106.

Piñango, Maria M, Jennifer Mack and Ray Jackendoff. 2006. Semantic combinatorial Processes in Argument Structure: Evidence from Light-Verbs. In *Proceedings of the Thirty-second Annual Meeting of the Berkeley Linguistics Society*.

Posner, M.I. and S.W. Keele. 1970. Retention of abstract ideas. *Journal of Experimental Psychology,* 83, 304-308.

Pradhan, Sameer, Eduard Hovy, Mitch Marcus, Martha Palmer, Lance Ramshaw and Ralph Weischedel. 2007. OntoNotes: A Unified Relational Semantic Representation, *First IEEE International Conference on Semantic Computing Proceedings*.

Prince, Ellen F. 1972. A Note on Aspect in English: the *take a walk* Construction. *Transformationelle Analyse* ed. by Senta Plötz, 409-40. Frankfurt: Athenäum.

Prins, A.A. 1952. *French Influence in English Phrasing*. Leiden: University of Leiden Publishing.

R Development Core Team. 2011. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rosch, E. and C.B. Mervis. 1975. Family Resemblances: Studies in the Internal Structures of Categories. In *Cognitive Psychology,* 7:4, 573-605.

Rosen, Sara. 1989. *Argument Structure and Complex Predicates*. Doctoral dissertation, Brandeis University.

Ross, Brian H., Susan J. Perkins and Patricia L. Tenpenny. 1990. Reminding-Based Category Learning. *Cognitive Psychology,* 22, 460-492.

Sag, I., Baldwin, T. Bond, F., Copestake, A., Flickinger, D. 2002. Multiword expressions: A pain in the neck for NLP. In Proceedings of the Third International Conference on Intelligent Text processing and Computational Linguistics (CICLING 2002), p. 1-15, Mexico City, Mexico.

Samardžić, Tanja and Paola Merlo. 2010. Cross-lingual variation of light verb constructions: using parallel corpora and automatic alignment for linguistic research. In Proceedings of the *2010 Worshop on NLP and Linguistics: Finding the Common Ground, ACL2010,* July 2010, pp. 52-60.

Sikos, L., Brown, S. W., Michaelis, L. A., and Palmer, M., & A. Kim. 2009. Figurative Language: Meaning is often more than just a sum of the parts. In *Proceedings of the 2008 AAAI Symposium on Biologically-Inspired Cognitive Architectures*, AAAI Technical Report FS-08-04. Menlo Park, CA: AAAI Press.

Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods,* 43(1): 155-167.

Stein, Gabriele. 1991. The Phrasal Verb Type 'to have a look' in Modern English. *International Review of Applied Linguistics in Language Teaching.* 29. 1-29.

Stevenson, Suzanne, Afsaneh Fazly and Ryan North. 2004. Statistical Measures of the Semi-Productivity of Light Verb Constructions. In Proceedings of the *Second ACL Workshop on Multiword Expressions: Integrating Processing,* July 2004, pp. 1-8.

Suttle, Laura and Adele E. Goldberg. 2011. Partial productivity of constructions as induction *Linguistics* 49 6: 1237-1269.

Tan, Y, M. Kan, and H. Cui. 2006. Extending corpus based identification of light verb constructions using a supervised learning framework. In *Proceedings of EACL-06 workshop on Multi-word-expressions in a multilingual context*, pp 49–56.

Tanabe, Harumi. 1999. Composite Predicates and Phrasal Verbs in The Paston Letters. In Brinton, L.J. and Akimoto, M. (Eds) *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*. Philadelphia: John Benjamins.

Theakston, Anna L. 2004. The role of entrenchment in children's and adults' performance on grammaticality judgment tasks. *Cognitive Development,* Volume 19, Issue 1, pp. 15-34.

Traugott, Elizabeth. 1999. A Historical Overview of Complex Predicate Types. In Brinton, L.J. and Akimoto, M. (Eds) *Collocational and Idiomatic Aspects of Composite Predicates in the History of English.* Philadelphia: John Benjamins.

Treiblmaier, H. and P. Filzmoser. 2009. Benefits from using continuous rating scales in online research. Forschungsbericht SM-2009-4.

Trueswell, J.C. & Kim, A.E. 1998. How to prune a garden-path by nipping it in the bud: Fast-priming of verb argument structures. *Journal of Memory and Language* 39, pp. 102-123.

Tu, Yuancheng, and Dan Roth. 2011. Learning English Light Verb Constructions: Contextual or Statistical. In *Proceedings of the Worshop on Multiword Expressions: from Parsing and Generation to the Real World (MWE 2011),* pp. 31-39.

Van Valin, Robert D., and LaPolla, Randy J. 1997. *Syntax: Structure, meaning and function.* New York: Cambridge University Press.

Venkatapathy, S and A. Joshi. 2005. Measuring the relative compositionality of verb-noun (v-n) collocations by integrating features. In *Proceedings of HLT and EMNLP05*, pp. 899–906.

Weischedel, R., E. Hovy, M. Palmer, M. Marcus, R. Belvin, S. Pradhan, L. Ramshaw and N. Xue. 2011. Ontonotes: A large training corpus for enhanced processing. In Olive, J.; Christianson, C.; and McCary, J., eds., *Handbook of Natural Language Processing and Machine Translation*. Springer.

Wierzbicka, Anna. 1972. *Semantic primitives*. Frankfurt: Athenäum.

Wierzbicka, Anna. 1980. *Lingua mentalis*. Sydney and New York: Academic Press.

Wierzbicka, Anna. 1982. Why can you Have a Drink when you can't *Have an Eat? Language*, 58(4):753–799.

Wilson, Stephen. 1999. *Coverbs and complex predicates in Wagiman*. Stanford, California: CSLI Publications.

Wittenberg, Eva & Maria M. Piñango. 2011. Processing Light Verb Constructions. *The Mental Lexicon* 6:3, 393–413

Wittenberg, Eva, Ray Jackendoff, Gina Kuperberg, Martin Paczynski, Jesse Snedeker and Heike Wiese. Submitted. The processing and representation of light verb constructions. http://www.wittenbergkuypers.de/Eva/publications_files/UncorrectedProofsWittenbegArgStr.pdf.

Xue,Nianwen, and Martha Palmer. Adding semantic roles to the Chinese Treebank. *Natural Language Engineering*. 15(1) 2009:143-172.

Zarco, M. A. 1999. Interlingual Representation of Complex Predicates in a Multilingual Approach: the Problem of Lexical Selection. Saint-Dizier, P. (Ed.) *Predicative Forms in Natural Language and In Lexical Knowledge Bases*. Dordecht: Kluwer Academic Publishers.

Zipf, George K. (1949) *Human Behavior and the Principle of Least Effort*. Addison-Wesley.