Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Kristin Denlinger 4/15/2014

The Role of Syntax in Word Conversion:

Uses and Limits of a Corpus-Based Approach to Converted Denominal Verbs

by

Kristin Denlinger

Phillip Wolff
Adviser

Program in Linguistics

Phillip Wolff

Adviser

Marjorie Pak

Committee Member

Benjamin Hary

Committee Member

2014

The Role of Syntax in Word Conversion:

Uses and Limits of a Corpus-Based Approach to Converted Denominal Verbs

By

Kristin Denlinger

Phillip Wolff

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Program in Linguistics

2014

Abstract

The Role of Syntax in Word Conversion:

Uses and Limits of a Corpus-Based Approach to Converted Denominal Verbs

By Kristin Denlinger

The current literature on denominal verb conversion has been dominated by subjective theories and has emphasized the semantic nature of the process. Hale and Keyser's theories of lexical noun incorporation and conflation uniquely suggest that syntax is a significant factor in the formation of denominal verbs. The present thesis aims to utilize empirical data to evaluate the legitimacy of these syntactic theories. The data originates from a syntactic parsing of Wikipedia, which includes information on frequency counts, types of verbal argument structures, and nominal roles of denominal verbs. Through the use of this data, it is argued that these theories can only be understood in the context of a dynamic and diachronic conversion process, in which semantic, syntactic and frequency factors of the converted pair are subject to change. The implications for this modified theory are discussed.

The Role of Syntax in Word Conversion:

Uses and Limits of a Corpus-Based Approach to Converted Denominal Verbs

By

Kristin Denlinger

Phillip Wolff

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Program in Linguistics

2014

Acknowledgements

I would like to thank Dr. Phillip Wolff for his support and mentorship during the development of this thesis. I would also like to thank all of the members of his Cognitive and Linguistic Systems Lab, particularly Austin Blodgett and Laura Manor for their crucial work in developing the corpus parsing program, Jason Shephard for his guidance and Hadar Naftalovich and Lee Kugelmann for their support and feedback. I would also like to thank Marjorie Pak and Benjamin Hary for their advice and guidance as well as for their participation as members of the committee for this thesis.

Table of Contents

List of Tables

List of Figures

Introduction

The present thesis aims to contribute to our understanding of converted denominal verbs, such as *crown* and *poison*, which have been defined in the literature as verbs that are formed directly from a noun base (Gottfurcht, 2008). Conversion is a word-formation process in which a word shifts lexical category without affixation.

Denominal verbs as well as deverbial nouns (nouns which are derived from verbs) are the most productive types of conversion in American English (Balteiro, 2007). In Balteiro's corpus analysis of converted lexical items, verbs that were derived from nouns, as opposed to "minor particles" (pg. 17) like adjectives, represented 93.10% of converted verbs, while deverbial nouns represented 97.23% of converted nouns. In other words, most converted nouns are derived from verbs and most converted verbs are derived from nouns (as opposed to adjectives, etc.).

There is a clear terminological distinction between deverbial nouns and denominal verbs. This distinction implies a uni-directional relationship between the parent and derived form, which I argue is not a useful distinction. Rather, I suggest that since all nouns and verbs have the potential to be converted to the other category, the process of conversion involves an activation and institutionalization of the derived form. Therefore, in recognition of the need for a directional-neutral term for these lexical units, I will henceforth refer to lexical items that can be used as verbs or nouns, regardless of their status as deverbial nouns or denominal verbs, as nerbs. For example, *table* is a nerb regardless of its use as a noun or a verb and regardless of the origin of the word. To refer to the noun use as well as the verb use, such as *table (n)/ table (v)* I will use the term nerb pairs.

The Present Thesis

The literature on denominal verbs is largely instinctive and theoretical. By utilizing empirical corpus data, we can make strides in answering questions where a subjective approach fails. While the literature on nerb conversion emphasizes the semantic processes involved, several aspects of conversion suggest that it involves a syntactic component. The syntactic roles of nouns and verbs are extremely useful when defining lexical categories and understanding lexical decomposition. Therefore, it seems intuitive that the process of conversion involves syntactic processes.

Hale and Keyser (1993; 2000) suggest two such models for a syntactically-based conversion of denominal verbs: movement-based noun incorporation and merge-like conflation. Both of these models successfully explain why certain morphemes can and cannot engage in conversion, questions that semantic-only theories are unable to answer. These theories are based on the assumption that the same syntactic principles that govern syntax at the overt level govern word-formation at the lexical level. Additionally, the syntactic ramifications of these word formation processes, such as traces, can have an effect on the use of the words after they are inserted into the syntax. Therefore it seems as though a syntactic theory of derivation will predict how nerbs behave at the overt level of syntax. However, this does not seem to be true. I will demonstrate using synchronic corpus data that nerb use does not suggest any syntactic categories that would indicate the type of derivation that they have undergone.

Although it seems on the surface that there are irreparable problems with syntactic theories of derivation, if we reanalyze our concept of derivation in the first place then these theories are still plausible. We cannot consider derivation to be a one-time operation in which the restrictions imposed by syntactic processes are fixed. Rather, it is important to understand

derivation as process of institutionalization that involves a series of innovations, in which both semantic and syntactic properties of the parent and derived words are subject to change. Therefore, what are traditionally termed derivational theories should instead be thought of as theories of only innovations because they are describing the first step in the process of conventionalization. In the present thesis, I argue that Hale and Keyser's syntactic theories can be considered valid only if we accept them as theories of *innovations* rather than *derivations,* and if any syntactic restrictions which were present at the time of innovation are subject to fade after lexical insertion. I will establish that several properties of a converted pair can change, including not only semantic change but also syntactic change, as the noun and verb frequencies shift.

Since the syntactic properties of nerb pairs are dynamic, I hypothesize that when a new verb innovation is converted from a noun, the verb will have very specific syntactic and semantic uses. Therefore, as the verb undergoes institutionalization and the use of the verb becomes more frequent, the syntactic restrictions on the verb will decrease. In order to empirically test how close a nerb pair is to the initial stages of innovation, I will use frequency counts which determine the proportional noun to verb use. In doing this, we can compare more "nouny" and "verby" nerbs with the variation of their argument structure. I predict that nounier verbs will be more syntactically restricted than verby verbs because nouny verbs are closer to innovations in the institutionalization process. Again, my hypothesis is that initial innovations are probably well predicted by syntactic theories. As will be discussed in greater detail below, certain syntactic theories of conversion hold that the head of a noun phrase is moved to the head of a verb phrase, a process that leaves a trace, and as has been amply demonstrated in the psycholinguistic literature, the presence of a trace places constraints on a verb's argument structure. However, when a nerb has an already established verb sense, syntactic features such as traces may be no

longer present in the representation of the nerb. Consequently, verby nerbs may be relatively unconstrained syntactically compared to nouny verbs. I will demonstrate empirically that as a nerb becomes nounier, the verbal use will be more restricted.

Denominal verbs are difficult to categorize syntactically from such synchronic data. However, our diachronic model of nerb conversion can still remain in compliance with Hale and Keyser's theories if we accept the idea that the restrictions imposed by syntactic models of innovations can fade as the verb becomes used more frequently and institutionalized.

Intuitions Supporting Syntax in Conversion

Kastovky (2005) notes that word-formation is "at the crossroads of morphology, syntax, semantic, pragmatics and the lexicon" (pg. 116). Marchand applies this concept to conversion specifically, saying that "only morphologically *and* semantically motivated combinations can give rise to new morphologically and semantically analysable formations" (as cited in Kastovky, 2005, pg. 101). While many of theories on conversion tend to focus on semantics, it is highly probable that both semantic and syntactic change occurs during the process of conversion.

Syntax in Lexical Categories

Conversion is characterized by a redistribution of semantic information across lexical categories. When we define lexical categories by their syntactic functions, particularly the syntactic functions of nouns and verbs, then it can be surmised that syntax plays a significant role in conversion. Mark Baker (2003) provides us with such a definition of lexical categories, one in which syntax is the most logical way to differentiate between nouns and verbs.

When it comes to defining lexical categories, traditionally minimalist frameworks have sought to break down nouns, verbs, adjectives and adpositions into binary distributions, often

grouping adjectives with nouns and grouping adpositions with verbs because of the types of argument structures in which they tend to appear (Baker, 2003). In Chomsky's (1970) original system, the four main grammatical categories could be distinguished in terms of the features +/- N and +/- V, with nouns being (+N, -V), verbs (-N, +V), adjectives (+N, +V) and appositives (-N, -V). Mark Baker (2003) expanded upon Chomsky's binary system, in holding that all four categories had a property of +-N/+-V. However, in Baker's analysis, the noun category (+N) is distinguished as a "referential index" (pg. 21) while the (+V) category is distinguished by its need for a specifier. Adjectives are defined by their lack of both of these features and prepositions are in a separate functional system altogether. Baker concludes that, ultimately, nouns and verbs are difficult to categorize by their semantic roles, but have clear differences in their syntactic functions. Therefore, there is an intuition that lexical categories add a syntactic component to the process of conversion in which the category changes.

Syntax in Lexical Decomposition

Rochelle Lieber describes conversion as "a sort of battle ground over which various theoretical camps have fought over the years" (Štekauer, 2005, pg. 418). The contention of these theoretical camps can be traced back to clashing understandings of word formation at the lexical level. For those who emphasize the role of semantics in conversion, the primary components of a word at the lexical level are semantic in nature, whereas syntactically based theories often require syntax to be relevant at the lexical level in order to explain word-formation processes. When words are governed by syntax at the lexical level, syntax becomes a plausible means through which word formation processes can occur.

For Hale and Keyser, this concept of syntax at the lexical level, termed l-syntax, as opposed to overt syntax, is critical to their intuition that syntax is a factor in word-formation. In

lexical semantics, it is widely accepted that lexical units have internal structure and can be broken down and understood as parts of a whole. In their discussion of the Lexicon-Syntax relation, Zubizarreta and Oh (2007) describe this common view: "The meaning of an expression is to be attributed to the superimposition of the meaning of grammatical closed-class items and the meanings of open-class items" (pg. 1). These closed-class items "correspond to universal concepts" that "are readily grammaticalised" (Pullman, 2005, pg. 5-6). This clearly assumes that words are not atomic, but rather vessels through which a semantically and syntactically restricted meaning can be utilized.

In l-syntax, the sub-atomic particles exist in a syntactic environment. Hale and Keyser break down verbs into smaller elements and all of these features give rise to syntactic relations. They claim that this system of l-syntax operates under the same syntactic processes and relations as words in the overt syntax (Pullman, 2005). I will use Pullman's example of the verb *kill* to illustrate this concept. The sentence *John killed Bill* can be decomposed to *John caused Bill to die*, in which the verb *die* raises by means of head to head movement to combine with the verb *cause* (see Figure 1 below).

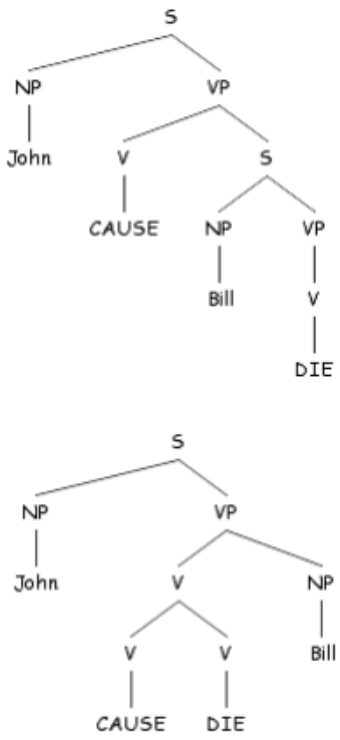

Figure 1. Word Formation in the l-syntax. Reprinted from Pullman, 2005.

Note that all of the elements are placed on syntactic heads and the fact that the inserted verb *kill* projects the NP *Bill* is the result of the causal nature of the original syntactic configuration. Therefore, the deep structure of the lexical item affects the syntactic behavior of the resulting verb. When new words are formed, the components of the decomposed verb are interacting syntactically. Hale and Keyser's theories on conversion will largely depend on the assumption that word formation processes are syntactically motivated.

Syntax in Categories of Denominal Verbs

The intuition that syntax plays a role in conversion is bolstered by the way in which we can intuitively classify denominal verbs. The most widely referenced and arguably the most comprehensive account of denominal verbs in a classification system comes from Clark and

Clark (1975). The authors divide denominal verbs into five classifications and a sixth miscellaneous category based on paraphrases which define each verb in terms of the noun from which it was supposedly derived. These categories are based on the authors' semantic interpretations and intuitions. For a detailed outline of these categories see Appendix. The authors admit that the paraphrases they use "are not themselves intended to carry any theoretical significance," as "most of the well-established verbs are specialized in ways not capturable in general paraphrases" (pg. 769). Therefore, Clark and Clark do not claim to have identified the sources from which the verbs have been derived. However, it is important to note that their categories are formed based on the role of the noun in the paraphrases, and they even note which grammatical case the parent noun takes. Therefore, while this classification is often referred to as a semantic classification, it is undeniable that syntax plays an important role in differentiating between these denominal verbs and, therefore, may play a role in the process of converting between them.

Conversion Independent of Syntax

While there are many intuitions that syntax is relevant to the conversion process, there are also prevalent theories which consider semantic processes to be more relevant, to the extent that they neglect syntax completely. Verbs and nouns in converted pairs are almost always clearly related. Consider the following examples and note the semantic similarities:

(1) to seat (v)/ a seat (n)

(2) to shelf (v)/ a shelf (n)

(3) to color (v)/ a color (n)

(4) to field (v)/ a field (n)

(5) to father (v)/ a father (n)

It is the semantic similarities between these pairs that cause linguists to assume a derivational relationship between the noun and the verb in the first place. Without this relationship, there would be no need for any theories about how denominal verbs or deverbial nouns are derived at all. We would assume that *seat* (v) and *seat* (n) are just as related to each other as *sofa* (n) and *lounge* (v) or *platypus* (n) and *curling* (v). Hale and Keyser (2000) note that their original interest in denominal verbs stemmed from this "denominal character" (pg.33) in which the nominal and verbal forms have identical phonological realizations and related meanings. Perhaps the clearest support for the existence of a nerb derivation process is the fact that verbs and nouns that are not in a nerb pair can easily be converted if necessary. For example, the noun *google* was quickly converted into a verb when English speakers had a need for a verb that, arguably, means *to use google to search for something.*

While semantics clearly play a large part in nerb conversion, there are several semantic theories that do not consider syntax to be involved in the process at all. Theories that neglect syntax often assume that syntax is not present at the lexical level at all and, therefore, at the word formation level. Accordingly, the important factor that influences whether syntax is useful in conversion depends on *where* we believe the derivation takes place in the word-formation process. For those theories that separate syntax from other word formation processes, the derivation takes place at a morphological, semantic or pragmatic phase.

For the Marchandeans, those who followed after the theories of Hans Marchand, conversion was not a particularly unique word formation process (Kastovsky, 2005). For Marchand, conversion was a process of zero-derivation, in which the stem is affixed as in any other derivation but the affix is phonologically null. This concept appears in many theories of derivation.

One of the most prominent theories that disregard syntax is the Lexicalist Hypothesis, in its many forms. This framework assumes the existence of a lexicon, an inventory of lexical and functional morphemes. In a strong lexicalist framework, word-formation is a process that occurs pre-syntactically in the lexicon. At the lexical level, derivations are morphological and follow a set of what Morris Halle deems "word formation rules" (as cited in Scalise & Guevara, 2005, pg. 166). Lapointe generalizes the lexicalist hypothesis as: "no syntactic rule can refer to elements of morphological structure" (as cited in Scalise & Guevara, 2005, pg. 170). A model of a simplified version of the lexicalist hypothesis is outlined below in Figure 2:

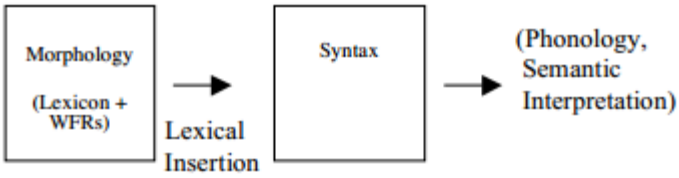

Figure 2 Lexicalist Word-Formation. Reprinted from Scalise & Guevara, 2005, pg. 173

Plag astutely demonstrates the perspective of semantically-focused theories when he states "that with any given productive affix, the syntactic category of potential base words is only a by-product of the semantics of the process" (as cited in Rainer, 1998, pg. 348). If we assume that the processes of conversion involve the use of an affix that is not phonologically realized, an assumption that in itself is highly controversial, then in these frameworks, conversion would just be another word formation rule. Therefore, syntax would not be involved in conversion at all, but would rather be a response to the semantic process.

While some proponents of the Lexicalist approach acknowledge that syntax plays at least a marginally important role, the onomasiological theory of word formation involves a completely

separate word formation component, which is related to the lexical component but is completely isolated from the syntactic component (Štekauer, 2005). A naming process, the onomasiological and onomatological levels, relate semantics and phonetics until the new lexeme is affixed in the lexicon and finally proceeds to the syntax.

These theories isolate syntactic processes from word-formation processes. However, there are certain questions that this perspective leaves unanswered. Particularly, these theories do not justify why certain morphemes cannot conflate, for example why nouns in the specifier position cannot conflate. Many of these questions are answered by the syntactic theories described below.

Syntactic Theories of Conversion

Movement-based Incorporation

Hale and Keyser's lexical noun incorporation and conflation theories follow the intuitions that syntax is at play in conversion, rejecting the above theories that neglect syntax. These theories generally follow in the tradition of Distributive Morphology, which is often pitted against the theory of Lexicalism. Distributive morphology, or DM, is a theory which rejects the notion of a lexicon in favor of an encyclopedia, which contains semantic information independent of phonological representations (Scalise & Guevara, 2005). In this framework, the phonetic realization of a word is not inserted until after syntactic heads interact, a principle called late insertion. The syntactic theories of derivation that follow can be understood to be assuming a more DM approach to word formation.

There have been several examples challenging the Lexicalist approach to word formation. The most essential to our present question is Baker's case of noun incorporation (1988), a syntactic word-formation process that involves "a syntactic movement of a word-level category

from its base position to combine with another word-level category" (as cited in Scalise &Guevara, 2005, pg. 176). Baker claims that "morphological derivations must directly reflect syntactic derivations (and vice versa)" (as cited in Scalise & Guevara, pg. 177).

Hale and Keyser (1993) apply Baker's noun incorporation in a syntactic theory of denominal verb formation. In this theory, the parent noun raises to the matrix verb via head-to-head movement at the lexical level. This movement is governed by the syntactic principles of Chomsky's *Government-Binding* framework, namely the *Empty Category Principle*, in which a trace must be properly governed (Chomsky, 1993).

Hale and Keyser validate many aspects of their theory by examining certain lexemes that are unable to undergo conversion in specific syntactic contexts. The most prominent of these syntactic limitations is the fact that denominal verbs "do not project a Specifier" (pg. 61). This means that if the parent noun is in the specifier position, it will not be able to incorporate into the verbal head because it is not properly c-commanded by the verbal head. Hale and Keyser (1993) claim that this is why we don't have verbs like *"*It machined the wine into bottles"* (pg. 60) to mean that a machine got the wine into bottles.

Similarly, the concept of minimality explains other ungrammatical sentences. In sentences like *"*He shelved the books on"* (pg. 60), the preposition is a barrier to the verb's proper government of the trace. Minimality also explains why an indirect object cannot be incorporated as in *"*She churched her money"* (pg. 60) in the sense that *she gave the church her money* because it would involve incorporation from the internal subject position. Additionally, they claim that only lexical heads can incorporate. Since prepositions are functional heads rather than lexical heads, they also cannot incorporate into the verb.

This theory of conversion is heavily dependent on the inclusion of syntax in our understanding of a word's lexical decomposition. Under this model, each denominal verb can be understood to have a deep structure, including a light verb and an argument that gets incorporated into the light verb. For example, the deep structure of the verb *shelve* includes the combination of the light verb *put* and the parent noun *shelf* in a prepositional phrase (Hale and Keyser, 1993).

Heidi Harley's (2005) analysis of telicity in denominal verbs within a DM framework supports Hale and Keyser's movement theory. She notes that the mass/count properties of nouns in locatum/location and unergative nerb pairs will affect the telicity of the resulting denominal verb. This finding suggests that there is a syntactic relationship between the parent noun and the derived verb because the type of noun that is being incorporated is affecting the syntactic restrictions of the resulting verb.

Manner Incorporation

While the incorporation theory of conversion responds to several questions about the syntactic restrictions of conversion, there are also a number of problems and limitations that the theory elicits. Harley (2005) points out a major limitation when she brings up different types of denominals. Hale and Keyser only address unergative, locatum and location verbs in their analysis. While Harley supports the movement-based theory for these types of verbs, it does not seem to work for what Clark and Clark (1979) have categorized as instrument verbs (in which the parent noun is in the instrumental case.) The reason for this is that the instrumental parent noun would be an adjunct rather than a complement in the deep structure and since the theory requires the parent noun to be a head in order to have head to head movement, the instrumental noun would be unable to incorporate into the verb head. This is clearly not the case, as nouns like

hammer can clearly be converted into verbs. Although Harley does not explicitly mention this, her problem of adjunct movement can also be applied to Clark and Clark's agent and experience verbs, as the paraphrase includes an adjunct, functioning almost like an adverb (i.e. *to butcher the cow* as *did to the cow the act that one would normally expect [a butcher to do to a cow]*, or *do in a butcher-like manner*).

Harley's (2005) explanation for this is that these types of verbs are derived in a different way from the incorporated verbs. She calls this process manner incorporation and describes it as "a mysterious, parametrically varying, ill-understood process" (pg. 3). In this process, "a *v* may be named by a Root describing the Manner in which it is accomplished" (pg. 26), as in Harley's model below (Figure 3.)

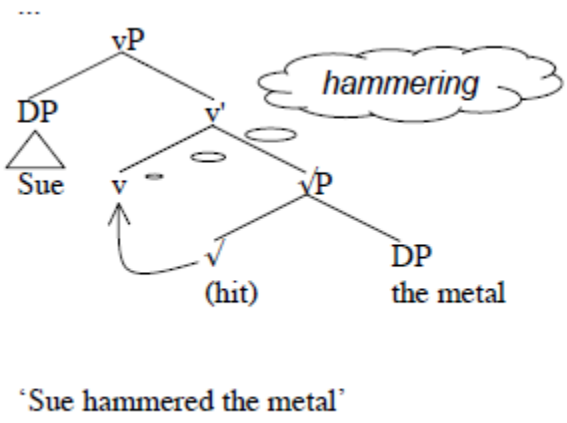

Figure 3. Manner Incorporation. Reprinted from Harley, 2005, pg. 26

Merge-like Conflation

Hale and Keyser also recognized flaws in their incorporation theory, and therefore proposed another theory in 2005 that mimicked the process of merge rather than move. The main flaw with the incorporation theory that Hale and Keyser recognize is that of cognate and hyponymous arguments. For example, the verb *laugh* can be used in the sentence *He laughed a*

hearty laugh. Also in the perfectly grammatical sentence *He shelved the books on the windowsill*, the argument *on the windowsill* is considered hyponymous (Hale & Keyser, 2000). According to the incorporation theory, these constructions would be impossible because the trace from the noun's movement would block these types of arguments. Therefore, as an alternative, Hale and Keyser propose a merge-based operation.

Merge is a process where two syntactic items are bound together (Hale and Keyser, 2000). Hale and Keyser call their merge-based theory conflation, a distinct process from the conflation that Clark and Clark discuss. In Hale and Keyser's conflation, the noun and verb are bound together and the p-signature, more or less the phonological feature of the nerb, is copied from one head into the other. In this model, the syntactic process is still executed before any phonetic insertion, but the nature of the process is very different because it doesn't result in a trace that needs to be properly governed. The p-signature is a property that only pertains to lexical items that are in the complement. Hale and Keyser use this concept to answer the question of why preopositions do not conflate, noting that prepositions have no p-signature since they are functional categories. Similarly, a verb cannot conflate with a noun in the specifier position because the noun does not have a p-signature since it is not in the complement.

New Implications for Syntactic Conversion Theories

The present study originally stemmed from a project examining the arguments taken by all verbs in English using a parsing of Wikipedia, a very large corpus of over 500 million words. The parsing program utilized the Stanford Parser as a basis. From this data, we were able to create matrices of syntactic information, including frequency counts for noun and verb use as well as a detailed account of the types of arguments taken by the verbs and a basic account of the

syntactic role of the noun. A more detailed explanation of the nature of these matrices will be explained where relevant in the following sections.

While there has been extensive theoretical work done on conversion processes, empirical basis for this research has been limited, particularly with a corpus of this size. It is important to note that the parsing of this data is entirely driven by syntactic structure and is entirely independent of semantics. This allows us to have a more objective and empirical look at argument structure and syntactic relations. The fact that semantic categories emerge from the parsing is a testament to the strength of the syntax-semantics interface. Regarding denominal verbs, by comparing nominal and verbal information from these parsings, we can obtain a clearer understanding of the syntactic relationship between nouns and verbs in nerb pairs with limited interference of semantic biases.

Categorization in Corpus Parsing

Through the use of semantic paraphrases, denominal verbs can easily be grouped into categories based on the nominal role of the parent noun in the paraphrases. This, as we recall, was the basis for Clark and Clark's categorization. Therefore, if derivation is in fact syntactic in nature, particularly if specific verbs undergo different types of syntactic processes such as noun incorporation versus manner incorporation, then we would predict that the syntactic behavior of the resulting denominal verbs reflect these differences.

In order to judge whether denominal verbs have similar syntactic behavior, we can plot the data from the verbal parsings in a visual space with respect to the syntactic similarities of the nerbs used. If different types of nerbs are derived through different types of syntactic processes, then we would expect the nerbs to cluster according to their syntactic tendencies. With our current data, this does not seem to be the case, as denominal verbs do not seem to cluster in groups or

even vary much in syntactic behavior, see Figures 4 and 5 below.

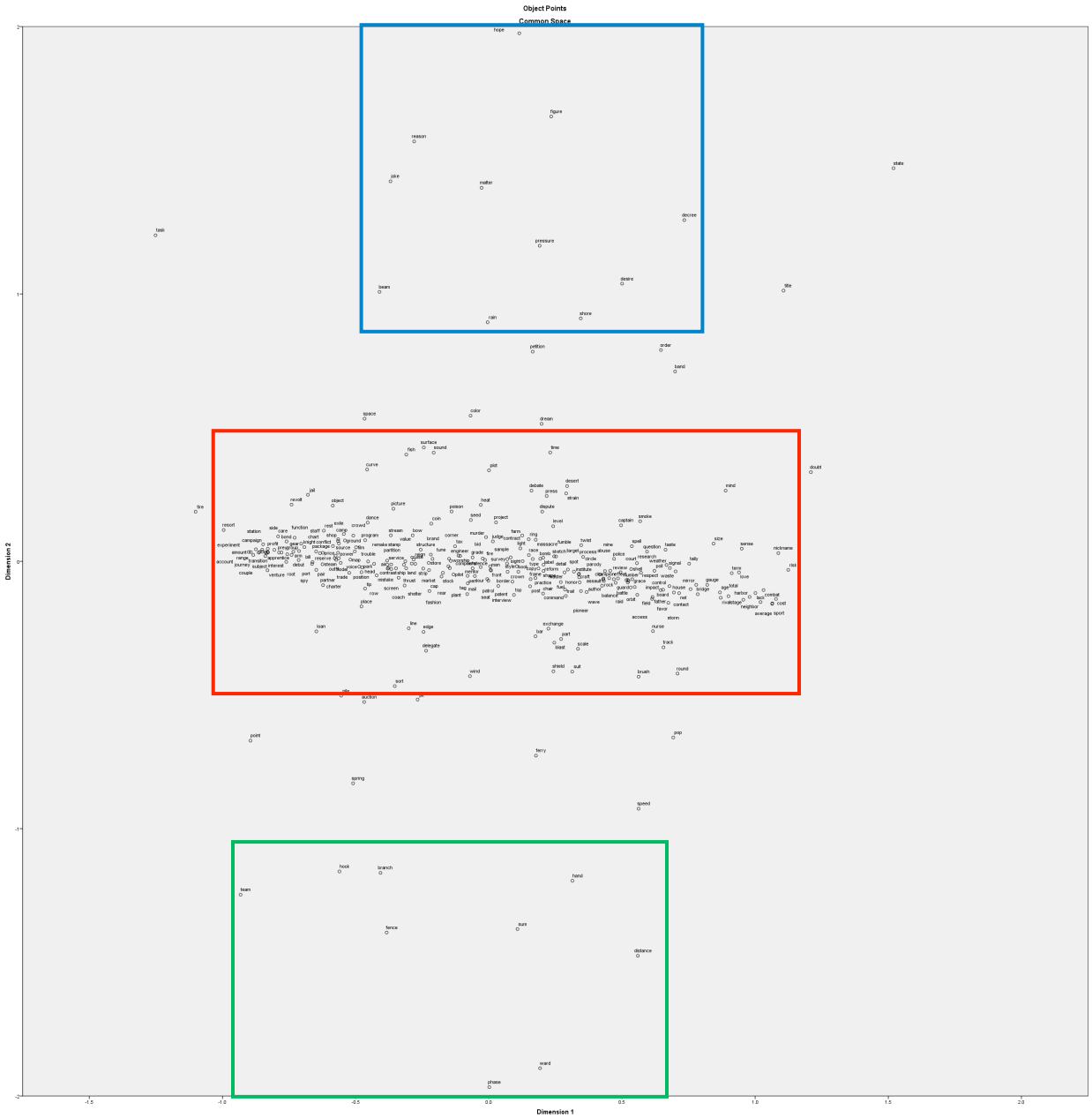

Figure 4. Denominals Plotted By Argument Structure

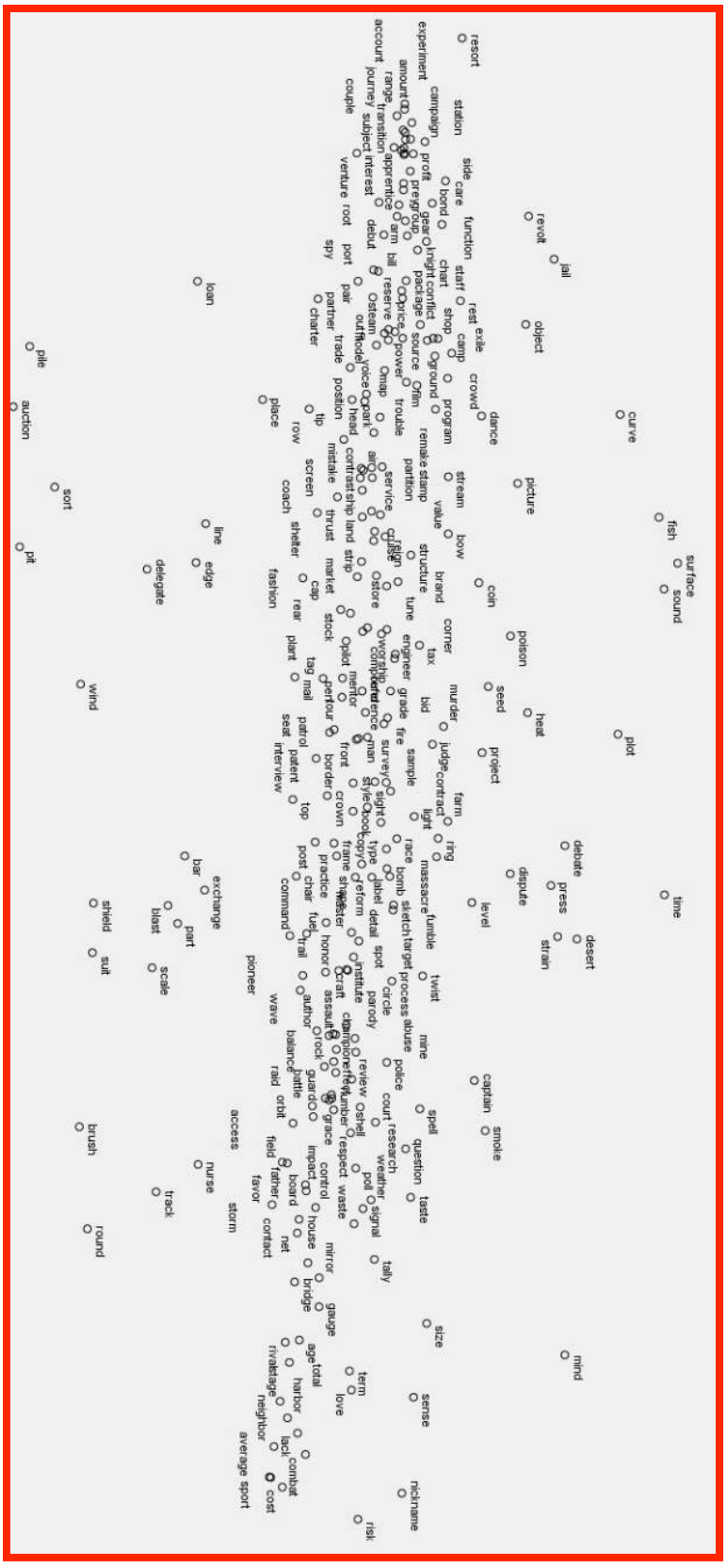

Figure 5. Center

Denominal Cluster

There are some odd denominals which have very restricted syntactic patterns, such as *hope* and *figure,* (see Figure 6 below) which can be used with S' arguments, such as *I hope that the defense goes well.* This structure does not seem to be characteristic of many denominal verbs and, therefore, these nerbs are plotted farther away from the general cluster of nerbs.

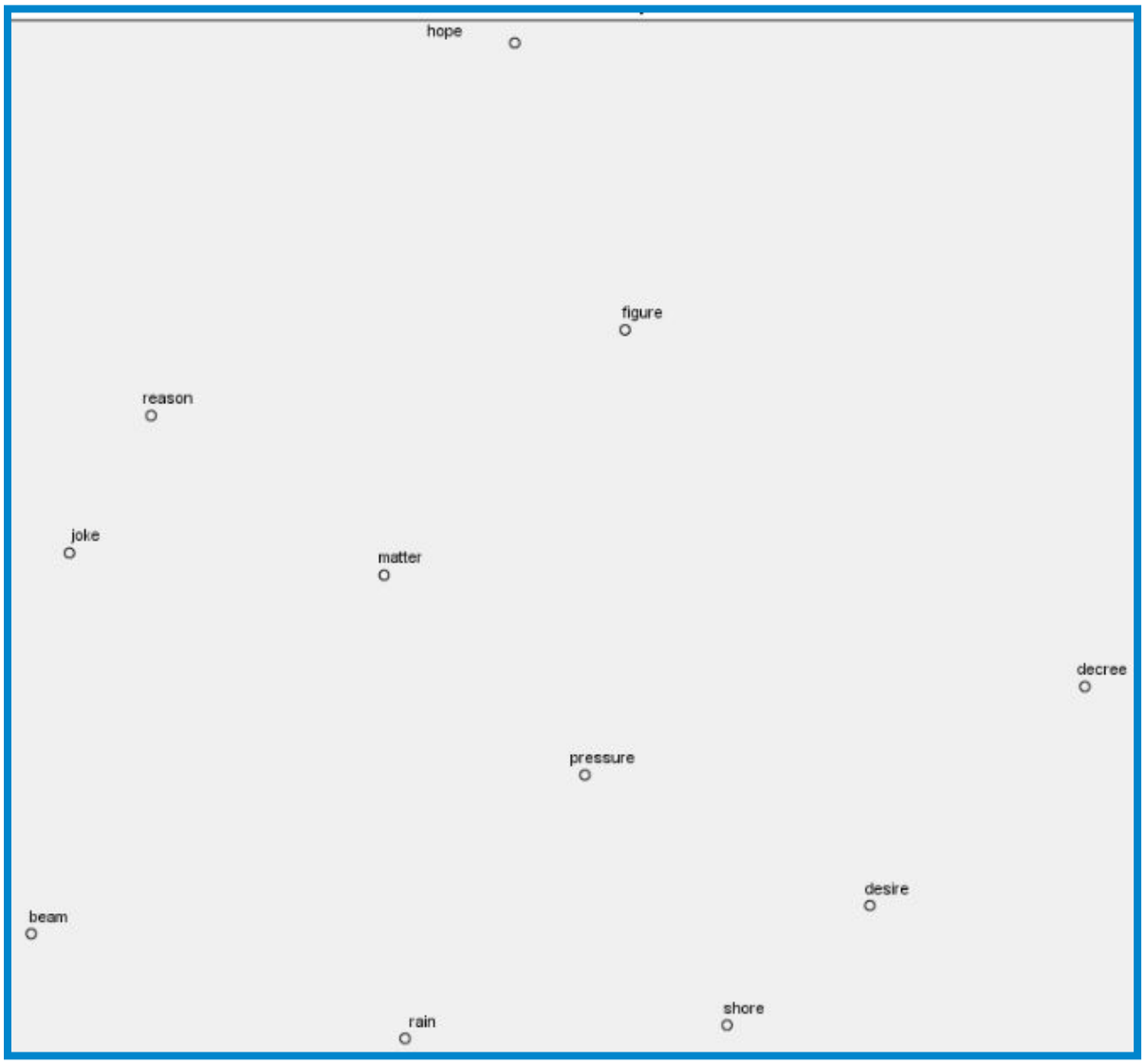

Figure 6. Outlying Denominals 1

On the other end of the space are verbs that occur with a specific preposition frequently (see Figure 7 below), such as *phase* as in *phase out* and *ward* as in *ward off.* These restrictions, however, are also sparse and the majority of the denominals cluster in a syntactically similar

space in the center of the plot.

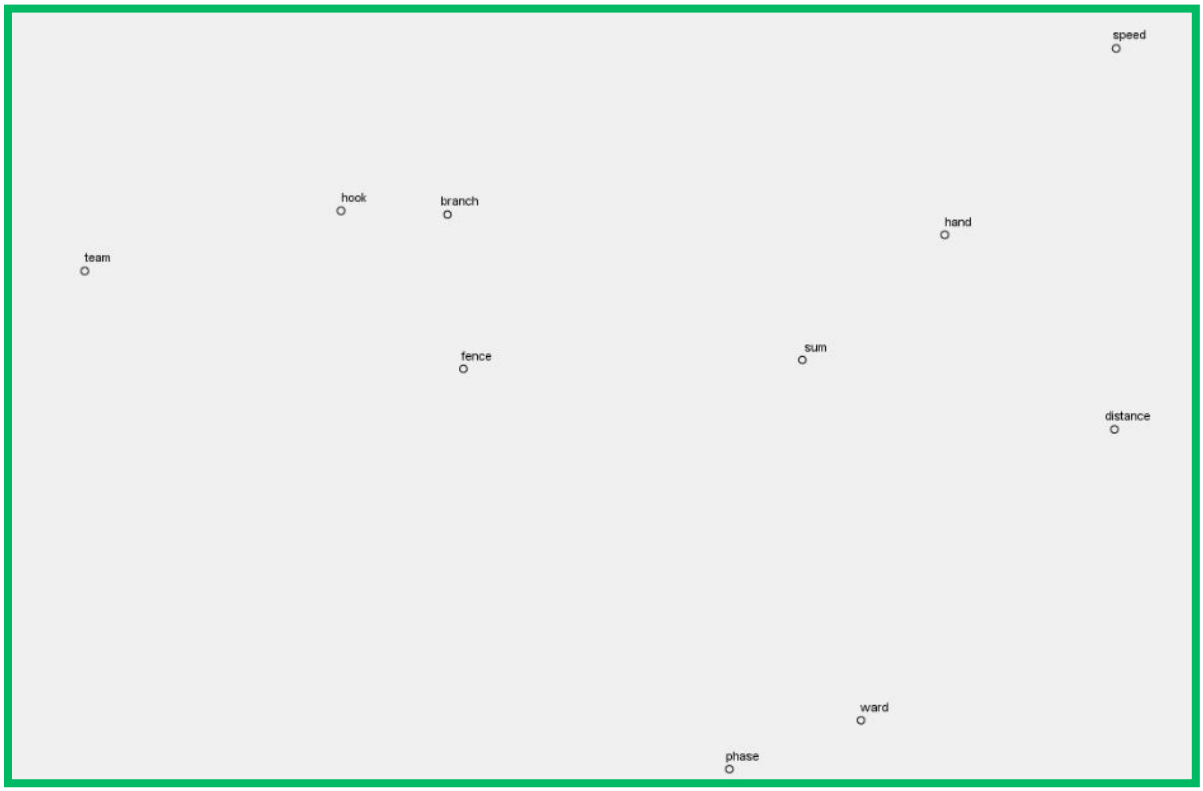

Figure 7. Outlying Denominals 2

Generally, this data suggests that there are not clear syntactic categories for denominal verbs, at least not based on the type of information that our parser accounts for. This does not necessarily mean that the derivational processes cannot be varied for different types of nerbs, but rather that the resulting denominals all behave in a syntactically similar way, or at least in a way that does not suggest distinct categories. If we believe that the derivational process is syntactic, then we would expect clear categories to emerge that differentiate between nerbs that have been derived through different syntactic processes.

Although lexical noun incorporation has been questioned and to some extent invalidated as an appropriate model for conversion, even by Hale and Keyser themselves, it can still provide a valid model if we reconsider the way in which we understand the conversion process.

Conversion as Innovation Rather Than Derivation

I argue that that above syntactic theories are still plausible when they are considered as theories of innovations rather than derivations. I will establish that syntactic restraints on nerbs, similar to the semantic uses of the nerb, are subject to change during the "institutionalization" (Kastovky, 2005, pg. 114) process, and justifying the fact that the synchronic data does not indicate categories of syntactic restraints. Clark and Clark (1979) make sure to emphasize the difference between innovations and derivations in the derivational process, acknowledging that these concepts exist on either end of a spectrum, and that it is difficult to draw a line between the two. The core of this question does not involve derivation at all, rather it asks how a new form becomes adopted and accepted into a language over time.

It has been argued that any noun can be converted into a denominal verb, as long as particular semantic conditions are followed when it is an innovation. Clark and Clark (1979) suggest that these conditions are pragmatic and include cooperation between speakers and contextual use. To use their example, if a boy in a group of friends was known to stroke people on the back of the knee with a teapot, the group of friends would be able to produce the innovation *teapoting* and would be able to mutually understand the phrase *Well, this time Max has gone too far. He tried to teapot a policeman,* to mean that Max stroked the back of a policeman's knee with a teapot. Therefore, under the right circumstances, any noun can be converted into a denominal verb as long as all speakers are able to understand the semantics of the innovation. This is the process of innovation rather than conversion. Therefore, we can say that the semantic relationships in converted pairs are the standardized and established result of a series of innovations.

Directionality Continuum as a Model for the Institutionalization Process

One of the relevant aspects of conversion which has interested linguists in the past is directionality. Directionality has been particularly controversial with respect to conversion since the process does not involve overt affixation and, therefore, it is difficult to tell which form is the parent word and which is the derived. Often, linguists have taken a uni-directional approach, assuming that one form must be derived from the other (Umbreit, 2010). These approaches ask whether the noun in a nerb pair has been derived from the verb or vice versa. Some (i.e. Balteiro, 2007) define directionality as a problem of which came first: the noun or the verb. Here, I will argue that directionality is best understood as a continuum rather than a dichotomy. While etymology is accurate in describing directionality as a singular innovation, more useful frequency measurement which identifies which form is more dominant and to what extent, is more useful in describing derivations, since it is a factor that can change over time.

Methods for diagnosing directionality in uni-directional approaches can be quite problematic. Hans Marchand was the first to propose a semantic definition of directionality. According to him, directionality is defined in terms of which form can be defined by the other (Balteiro, 2007). However, one can easily argue against such a definition of directionality. For example, *to hammer* can be defined as *to use a hammer*. However, *a hammer* can also be defined as *the tool with which one hammers.* Given that either the verb or noun can be defined in terms of the other, such an approach to directionality does not resolve the question of which form is most basic. Essentially, these arguments are subjective and cyclical, and the fact that verbalization and nominalization both exist in the first place suggests that either direction is possible Other theories of semantic directionality suggest the use of metonym in determining

which use is more "well-entrenched" (Umbreit, 2010, pg. 311). However, these interpretations fall victim to a similar subjectivity flaw, as the semantic basic-ness of a certain form is debatable.

Perhaps the most obvious strategy for determining directionality is the use of etymological evidence from the first uses of the nerb in each category. Balteiro (2007) considers this tactic to be the most telling; she argues that whichever use appeared first must be the parent unit. However, etymology can be problematic when interpreting synchronic data because it assumes that the origin of the derivation accurately corresponds to the present use of the nerb pair. Although the nerb often maintains the parent and derived relationship that was existent at the time of the derivation, it is also very possible that the nerb use now is very different from the nerb use at the time of innovations.

The last criterion for assessing directionality is frequency, the most pertinent to our corpus-based study. In this approach, whichever use is more frequent is considered the parent use and the less frequent use is the derived form. For example, the derived word *blackbird* will probably be used less frequently than either *black* or *bird.* It might be more apt to say that the more frequent use is the more dominant use, and is therefore cognitively considered to be the parent use.

While the frequency method has been used in uni-directional approaches, I argue that frequency is useful for the synchronic corpus because it describes the current cognitive balance of the noun and verb form of nerbs. Additionally, it is a factor that can change over time. This flexibility is key if we consider derivation to be a series of innovations over time. In my analysis, I will refer to this factor as Nouniness, which is calculated as the number of nominal uses divided by the total number of uses. Therefore, a nerb with significantly more noun uses than verb uses

will be closer to 1, but a nerb with more verb uses will be closer to -1. A nerb whose noun and verb usage is close to balanced will have a Nouniness rating close to 0.

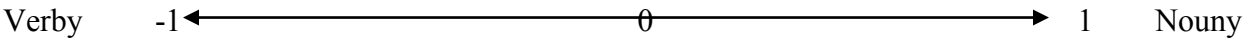

Figure 8. The Nouniness Spectrum

Nouniness is not necessarily a measurement of directionality, per se, as long as directionality is defined as a dichotomous judgment of which form is the parent noun. However, if we understand directionality as a continuum, then Nouniness becomes more useful. Nouniness takes into account the changes in balance between nouns and verbs that can occur since the first uses in either category.

Turning to evidence from the corpus parsing, it is clear that etymology does not necessarily predict Nouniness. Let us first consider the nerbs in Table 1 below, sampled from the ends and middle of the Nouniness spectrum, in which etymology corresponds with the current frequency usage of the nerb. For each nerb, the earliest noun use and earliest verb use according to the Oxford English Dictionary (OED) are provided. The type of use which occurred first is highlighted in green. Additionally, the difference in years between the earliest use and the "novel" use is provided. This measure is helpful because it suggests whether the earlier use was well-established before the derivation occurred. For nerbs in which the derivation occurred within 10 years of the earliest use, the cells are colored a lighter green. For nerbs which had the earliest noun and verb use occur in the same year, the cells are colored yellow. Finally, a measure of Nouniness is provided according to the frequency counts from the corpus parsing.

Table 1. Etymology Predicting Nouniness in a Selected Sample

Nerb	Earliest Noun Use (year)	Earliest Verb Use (year)	Difference in years	Nouniness
Make	1325	pre 1150*	at least 175	-0.99989
Say	1571	1000	571	-0.99909
Sell	1952	1000	952	-0.99907
Think	1834	pre 1150*	at least 684	-0.99671
Leave	1513	1225	288	-0.99298
disgust	1598	1601	3	-0.42266
delight	1225	1225	0	-0.416
Dream	1300	1300	0	-0.33158
Joke	1670	1670	0	-0.305
Sound	1297	1300	3	-0.304
Jab	1825	1825	0	-0.0297
Seesaw	1704	1712	8	0
Taste	1340	1340	0	0.004
Mutiny	1567	1584	17	0.007453
Father	825	1483	658	0.940267
School	pre 1150*	1456	at least 306	0.991042
Baby	1400	1744	344	0.9923
Throne	1225	1377	152	0.99363
summer	825	1440	615	0.982

**originated during Old English (OE) period, calculated from the end of OE, the year 1150.*

Obviously, in these cases the determination of etymologically-based directionality predicts Nouniness. In other words, the direction and distance between the verb and noun uses predicts how balanced our contemporary uses are. The relationship between the nominal form and the verbal form has not changed since its earliest uses. However, this not the case for all nerbs.

Consider, for example, the set of nerbs in chart 2 :

Table 2. Etymology Unable to Predict Nouniness in a Selected Sample

Nerb	Earliest Noun Use (year)	Earliest Verb Use (year)	Difference in years	Nouniness
Feud	1400	1673	273	0.01718
Hike	1865	1809	56	0.027
Poison	1225	1350	125	-0.02061
Nap	1400	pre 1150	250	-0.299
Review	1441	1573	132	-0.296
Drink	888	1000	112	-0.7624

For these nerbs, etymology does not correspond with the current frequency of noun and verb uses. For *feud* and *poison* the nominal use occurred well before the first verbal use, and yet at the time of the parsing both forms were balanced in use. Conversely, *hike* and *nap* were well-established verbs before the earliest noun use appeared, and yet their Nouniness indicates that they are closer to the balanced point of the spectrum. *Review* historically began as a noun, and not only has its Nouniness moved closer to a balanced rating, but it has even reversed the frequency, making the verbal use slightly more frequent than the nominal use. *Drink* provides an

even more drastic transition, as it originated as a noun but has a fairly verbal Nouniness rating. Therefore, etymological derivation does not necessarily predict the relationship between the noun and the verb use. This is important because it indicates that while etymology may provide a rating of dichotomous directionality, it fails to take into account the shifting frequencies in our usage over time, which arguably continues to change the cognitive status of a nerb.

Semantic and Syntactic Change over Time

It is clear that nerb pairs are subject to semantic change over time. For a nerb like *seat,* both the nominal and the verbal forms of the nerb seem to be fairly concrete and referential to a specific concept or event. However, a nerb like *ground* has a greater discrepancy between its noun and verb uses that make the derivative relationship less clear. While the nominal version of *ground* often refers to a fairly concrete and referential concept, the verb form of *ground* can be applied to a more figurative uses, such as *ground your thoughts, ground the argument, etc.* Of course, the nominal version can be used figuratively, as in something like *the grounds for early termination,* and the verbal version can be used more concretely, as in *ground the airplane*, and yet each form tends to be used in different ways. We can hypothesize that, in the early stages of a nerb derivation, that is during the initial innovations, the use was more concrete and as the derivation became staler the use was extended metaphorically. This phenomenon is an interesting example of how a nerb pair can evolve semantically, and presents the danger in using synchronic data without considering how the nerb pair can change over time.

Seeing as Nouniness is a factor that can also change over time, and if syntax is at play during the conversion process, then we would also expect syntactic change to occur to a nerb pair over time. I predict that, just as an innovated form tends to be more semantically specific than its parent form, an innovated nerb will also have syntactic restrictions. Hence, in an

unbalanced nerb pair the less frequent form will have more syntactic restrictions than the more frequent form. For example, the very "verby" nerb *leave* is fairly unrestricted in its verbal form, and yet the nominal form, as in *give leave* or *take leave*, is fairly restricted in the sense that it can only occur as the direct object of very specific verbs. On the other hand, the very "nouny" nerb *father* can be used in a variety of nominal positions, yet the verbal form is almost always followed by an NP direct object.

Consequently, we expect that as a nerb becomes verbier, the nominal form becomes more restricted. Conversely, as a nerb becomes nounier, the verbal form becomes more restricted. In other words, as Nouniness increases, nominal constructions are less restricted and verbal constructions are more restricted. This hypothesis can be tested with the information that we can glean from our corpus matrices.

The verbal matrix has information on frequency, and therefore Nouniness, in addition to the types of arguments that each verb takes and the number of times that it takes each argument. The parser counts up to three arguments after each verb. For example, in the sentence *The man risked his life for his money*, the verb *risk* takes two arguments, the NP (noun phrase) *his life* followed by the PP (prepositional phrase) *for his money*. The third argument is null, therefore this instance of *risk* would warrant the arguments NP PP 0. The parser also accounts for the order of the arguments, for example the NP PP 0 above is different from the arguments PP NP 0. The parser also differentiates between prepositions, for example PP headed by *for* versus a PP headed by *in.* Although the parser uncovered around five thousand combinations of arguments, the matrix used only considered the top one hundred constructions since the less frequent arguments after one hundred became extremely rare. In our verbal matrices, we can measure restriction by the variation of the arguments that the verb is able to take. However, this method falls victim to

frequency effects, in which extremely frequent verbs may seem more varied than extremely low count verbs which only have the opportunity to occur in a low number of arguments due to their size. In order to account for this, we used a random sample in which only the first five hundred instances of each were considered.

I predicted that this verbal variation measurement will decrease as Nouniness increases. This prediction turns out to be accurate. Nouniness and argument variation are negatively correlated ($r= -.147$, $n=2068$, $p<.01$).

This correlation supports the notion that denominal verbs that are innovated have more restricted syntactic roles, but those restrictions can fade over time as the new verb form becomes more frequently used. If nerbs have a tendency to move towards a more balanced state between their noun and verb uses, since the novel use must be productive to some degree in order to survive, then Nouniness can be used to determine whether the nerb is closer to the innovation end or the well-established end. It makes sense then that nouny nerbs will have slightly more restricted syntax.

Implications for Syntactic Theories

If we accept this paradigm that derivation involves a series of innovations and that semantic and syntactic properties are flexible as the frequency of noun and verb uses change, then syntactic theories are seemingly problematic. Semantic theories of derivation are more forgiving of semantic changes then syntactic theories are of syntactic changes. Semantic properties are more subjective and less systematic. Recall Clark and Clark's claim that their semantic paraphrases are specialized beyond their general paraphrases. This idea in and of itself implies that the semantic relationship between the parent and derived forms is loose. However, syntactic properties result in more strict and fixed transformations, such as traces. Hale and

Keyser's l-syntax operates under the same principles as overt syntax so these syntactic restrains affect lexemes during and after the word formation process. With this understanding, we would not expect the syntactic properties to change over time.

I argue that this is not the case. In order for syntactic theories to be applicable under our diachronic understanding of nerb derivation, we must accept the fact that the lexical syntax properties that govern word formation processes, such as conversion, under Hale and Keyser's model can fade after lexical insertion. Therefore, the problematic traces from syntactic movement which would prevent hyponymous arguments in the overt syntax are no longer relevant in the overt syntactic level. So, the noun-incorporation movement theory, coupled with Heidi Harley's manner incorporation theory, would still be plausible at the lexical level, but the resulting syntactic effects of the movement would not be permanent after lexical insertion.

Similarly, the conflation model is also plausible under the diachronic view of derivation. Although merge processes do not result in traces that need to be governed like movement processes do, it is still important to consider the conflation model in the context of dynamic nerbs. If we interpret the theory of conflation through this lens, we must include Nouniness as a factor during the process of conflation.

The "mergability" or "conflatability" of the noun and verb during conflation is already discerning in nature. That is to say, only certain nouns can undergo conflation depending on their syntactic role. If that is so, and if we look at derivation as a process that occurs many times before a nerb is well-established, then during conflation there may also be a Nouniness factor which influences the likelihood of the nerb's conflation. For example, if a noun has an extremely high measure of Nouniness, then it will be less likely to conflate with the verb. If, over time, the verbal form becomes more established and the frequency becomes more balanced, then the noun

will more easily conflate to the verbal category. This theory allows conflation to be a more dynamic process that takes into account the directionality spectrum.

Conclusions and Future Research

The previous thesis has argued that syntactic theories of nerb conversion are still feasible as theories of innovation rather than theories of derivation. Since the collected data suggests that syntactic constraints can dissolve during the process of institutionalization, it is still highly possibly that syntactic properties govern innovations. However, the syntactic restrictions of lexical syntax can fade after the verb is inserted into the overt syntax.

Many questions remain regarding the nature of denominal verb derivation and the role that syntax plays in these processes. It is clear that this process is at the intersection of nearly every aspect of linguistics, particularly semantics and syntax. However, with a directionality continuum and a dynamic understanding of nerb properties, it seems as if syntax plays a notable role in the formation of converted nerbs.

As with all corpus studies, it would be useful to attempt to duplicate these findings with data from another corpus. While many projects could stem from this research, one next step could be the search for a more empirical approach to determining the paraphrases of denominal verbs, particularly identifying which light verbs each nerb paraphrase includes. This could provide more information on the role of syntax in the derivational process because different light verbs project different syntactic structures.

One of the most problematic aspects of utilizing corpus data to relate semantically connected nerb pairs is that it does not account for polysemy. Admittedly, this is a problem. However, considering the limitation of corpus parsings with respect to parsing multiple meanings, it is a particularly massive obstacle for corpus-based analyses. For the present study,

the practical solution to this problem is not worth pursuing considering the size of the corpus. I argue that, although the polysemy problem will partially confound the data, a frequency measure for a corpus of this size has still provided us with a useful quality of the data. Perhaps in future research, particularly with smaller corpuses, polysemy could be accounted for to provide a more accurate account of verb and noun use.

In this thesis, Nouniness is used as a measure of how far along the nerb is in the process of institutionalization. This is not the ideal measure, since other factors can influence a nerb's transition from innovation to conventionalized form. For example, it may be true that verbs which are derived from more frequent nouns enjoy a speedier transition towards institutionalization. Factors such as these could be incorporated into the Nouniness measures to glean a more accurate measure of a nerb's institutionalization status. These measures could be beneficial to our understanding of nerb derivation in future research.

References

Primary Sources

Dictionary, O. E. (1989). Oxford: Oxford University Press.

Wikipedia: The free encyclopedia. (2004, July 22). FL: Wikimedia Foundation, Inc. Retrieved August 10, 2004, from http://www.wikipedia.org

Secondary Sources

Baker, M. C. (1988). Incorporation: A theory of grammatical function changing.

Baker, M. C. (2003). *Lexical categories: Verbs, nouns and adjectives* (Vol. 102). Cambridge University Press.

Balteiro, I. (2007). *The directionality of conversion in English: a dia-synchronic study* (Vol. 59). Peter Lang.

Chomsky, N. (1970). Remarks on nominalization. In *Readings in English transformational grammar*,ed. R. Jacobs and P. Rosenbaum, 184–221. Waltham, Mass.: Ginn.

Chomsky, N. (1993). *Lectures on government and binding: The Pisa lectures*(No. 9). Walter de Gruyter.

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, 767-811.

Gottfurcht, C. A. (2007). *Denominal verb formation in English*. ProQuest.

Hale, K., & Keyser, J. (2000). Conflation.

Hale, K., & Keyser, J. (2013). Some transitivity alternations in English. *Anuario del Seminario de Filología Vasca" Julio de Urquijo"*, *20*(3), 605-638.

Hale, K., & Keyser, S. J. (1993). On argument structure and the lexical expression of syntactic relations. *The view from Building*, *20*(53-109).

Harley, H. (2005). How do verbs get their names? Denominal verbs, manner incorporation, and the ontology of verb roots in English. In *The syntax of aspect*.

Kastovsky, D. (2005). Hans Marchand and the Marchandeans. In *Handbook of word-formation* (pp. 99-124). Springer Netherlands.

Klein, D. & Manning, C. D. (2003). Fast Exact Interference with a Factored Model for Natural Language Parsing. In *Advances in Neural Information Processing Systems 15 (NIPS, 2002).* Cambridge, MA: MIT Press, pp. 3-10.

Lieber, R. (2005). English word-formation processes. In *Handbook of word-formation* (pp. 375-427). Springer Netherlands.

Marchand, H. (1966). *The categories and types of present-day English word-formation: A synchronic-diachronic approach* (No. 13). University of Alabama Press.

Pulman, S. G. (2005). Lexical decomposition: For and against. In *Charting a New Course: Natural Language Processing and Information Retrieval* (pp. 155-173). Springer Netherlands.

Rainer, F. (2005). Constraints on productivity. In *Handbook of word-formation*(pp. 335-352). Springer Netherlands.

Scalise, S., & Guevara, E. (2005). The lexicalist approach to word-formation and the notion of the lexicon. In *Handbook of word-formation* (pp. 147-187). Springer Netherlands.

Štekauer, P. (2005). Onomasiological approach to word-formation. In *Handbook of word-formation* (pp. 207-232). Springer Netherlands.

Tuggy, D. (2005). Cognitive approach to word-formation. In *Handbook of Word-Formation* (pp. 233-265). Springer Netherlands.

Umbreit, B. (2010). Does Love Come From to Love or to Love From Love? Why Lexical Motivation Has to Be Regarded as Bidirectional. *Cognitive Perspectives on Word Formation*, *221*, 301.

Zubizarreta, M. L., & Oh, E. (2007). *On the syntactic composition of manner and motion*. MIT Press.

Appendix

Outline of Clark and Clark's (1975) nerb categories:

The first category contains locatum verbs, which include a causative component and the parent noun in a prepositional phrase. The parent noun is in the objective case. For example, the verb *blanket* as in *Jane blanketed the bed* can be paraphrased to "*Jane did something to cause it to come about that [the bed had one or more blankets on it*]" (Clark & Clark, 1975, pg. 769). Note that the parent noun is what is being put on something else (the bed).

The second category contains location and duration verbs, which also include a causative component and the parent noun in a prepositional phrase. However, the role of the parent noun in the prepositional phrase is the opposite of the parent noun role in locatum verbs. In other words, the parent noun is in the locative case. For example, the verb *kennel* as in *Kenneth kenneled the dog*, can be paraphrased to "*Kenneth did something to cause it to come about that [the dog was in the kennel]"* (Clark and Clark, 1975, pg. 772). In this case, the parent noun, *kennel,* is the thing in which the dog is being put rather than the thing that is being put in something.

The next category contains agent and experience verbs, which involve doing something in a way that is expected of the parent noun, which is in the agentive case. For example, the verb *to butcher* as in *John butchered the cow* can be paraphrased to "*John did to the cow the act that one would normally expect [a butcher to do to a cow]"* (Clark and Clark, 1975, pg. 773).

The next category includes goal and source verbs, in which the paraphrase also contains a causative component but the parent noun is involved in describing the goal or source of the causal event. For example, the verb *powder* as in *Edward powdered the aspirin* can be

paraphrased to "*Edward did something to cause it to come about that [the aspirin was powder]"* (Clark and Clark, 1975, pg. 774).

Lastly, they identify instrument denominals, in which the parent noun is an instrument involved in a causative event. For example, the verb *bicycle* as in *John bicycled into town* can be paraphrased as "*John caused it to come about that he was in town by doing the act one would normally expect [one to do with a bicycle]"* (Clark and Clark, 1975, pg. 776). Note that this paraphrase is similar to that of the agent and experience verbs, but it involves a causative component.