

Fall 11-10-2016

# Lexical variation, lexical innovation, and speaker motivations: a historical psycholinguistic approach

Jason Timm Dr.  
*University of New Mexico*

Follow this and additional works at: [https://digitalrepository.unm.edu/ling\\_etds](https://digitalrepository.unm.edu/ling_etds)



Part of the [Computational Linguistics Commons](#), and the [Discourse and Text Linguistics Commons](#)

---

## Recommended Citation

Timm, Jason Dr. "Lexical variation, lexical innovation, and speaker motivations: a historical psycholinguistic approach." (2016).  
[https://digitalrepository.unm.edu/ling\\_etds/45](https://digitalrepository.unm.edu/ling_etds/45)

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Linguistics ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact [disc@unm.edu](mailto:disc@unm.edu).

**Jason Timm**

*Candidate*

---

**Linguistics**

*Department*

---

This dissertation is approved, and it is acceptable in quality and form for publication:

*Approved by the Dissertation Committee:*

William Croft, Chairperson

---

Jill Morford

---

Christian Koops

---

Richard File-Muriel

---

---

---

---

---

---

---

---

**LEXICAL VARIATION, LEXICAL INNOVATION, AND SPEAKER  
MOTIVATIONS: A HISTORICAL PSYCHOLINGUISTIC APPROACH**

by

**JASON TIMM**

B.S., Environmental Systems Technology, Cornell University, 2002  
M.A., Linguistics, University of New Mexico, 2010

DISSERTATION

Submitted in Partial Fulfillment of the  
Requirements for the Degree of

**Doctor of Philosophy Linguistics**

The University of New Mexico  
Albuquerque, New Mexico

**December 2016**

**LEXICAL VARIATION, LEXICAL INNOVATION, AND SPEAKER  
MOTIVATIONS: A HISTORICAL PSYCHOLINGUISTIC APPROACH**

**By**

**JASON TIMM**

B.S., Environmental Systems Technology, Cornell University, 2002

M.A., Linguistics, University of New Mexico, 2010

Ph.D., Linguistics, University of New Mexico, 2016

**ABSTRACT**

Speakers commonly re-purpose existing forms in the mental lexicon to create novel form-meaning. Contemporary evidence that such innovation processes have occurred historically is attested in varying degrees of polysemy in the mental lexicon. This dissertation considers speaker motivations underlying these innovation processes historically. Strong synchronic relationships between frequency and degree of polysemy, on one hand, and frequency and lexical access, on the other hand, have traditionally been interpreted as evidence for the primacy of economic motivations in processes of lexical innovation. In contrast, the cognitive processes that most commonly facilitate innovation, metaphor and metonymy, have largely been described as processes motivated by expressiveness and not being misunderstood.

In order to assess the role of these competing motivations in processes of innovation, an idealized model of lexical change is presented, in which the corpus-distributional characteristics of forms used in innovation synchronically are considered independently from (1) the characteristics of polysemous forms synchronically, and (2) the distributional and diachronic consequences of the propagation of novel form-meaning in the speech community. Based on the corpus-distributional characteristics and degree of polysemy of approximately 20 thousand word forms in American English at three

synchronic “points” in time (1810-1849, 1910-1929, and 1990-2009), three historical models are presented: a model of polysemy, a model of lexical innovation, and a model of propagation.

Results from the model of innovation demonstrate evidence of competing motivations (ie, economy, expressiveness, and not being misunderstood) in processes of innovation historically; importantly, model results demonstrate that the synchronic corpus-distributional characteristics of forms used in innovation are indeed distinct from the synchronic characteristics of polysemous word forms as described in the model of polysemy. Results from the model of propagation demonstrate the distributional consequences of successful propagation historically, and provide further evidence for the role of competing motivations historically. In sum, the studies presented in this dissertation demonstrate important roles for the speaker motivations underlying processes of metaphor and metonymy that have generally been overlooked by strictly synchronic approaches to lexical innovation.

## TABLE OF CONTENTS

<b>LIST OF FIGURES.....</b>	<b>viii</b>
<b>LIST OF TABLES.....</b>	<b>ix</b>
<b>CHAPTER 1. INTRODUCTION.....</b>	<b>1</b>
§1.1 Identification of the problem .....	1
§1.2 Processes underlying lexical innovation.....	2
§1.3 Speaker motivations.....	4
§1.3.1 Economy.....	5
§1.3.2 Expressiveness and not being misunderstood .....	8
§1.3.3 Competing motivations.....	10
§1.4 Speaker and speech community methodologically.....	12
§1.5 Dissertation organization.....	14
<b>CHAPTER 2. POLYSEMY HISTORICALLY.....</b>	<b>17</b>
§2.1 Methodological approaches to degree of polysemy and lexical change .....	17
§2.1.1 Corpus-based approaches .....	17
§2.1.2 Corpus-external approaches .....	19
§2.1.3 Arguments for a corpus-external approach.....	21
§2.2 An historical dictionary methodology .....	21
§2.3 Evidence of lexicographic consistency historically.....	25
§2.4 Evidence of lexical development historically.....	28
<b>CHAPTER 3. POLYSEMY IN THE SPEECH COMMUNITY.....</b>	<b>32</b>
§3.1 Aligning dictionaries and speech communities historically .....	32
§3.2 Distributional features of lexical variation .....	35

§3.2.1 Dissemination and frequency .....	36
§3.2.2 Discourse environments .....	39
§3.2.3 Local environments.....	42
§3.2.4 Word length in syllables and part of speech.....	46
§3.2.5 Summary: Features of distributional variation .....	47
§3.3 A model of polysemy.....	49
§3.3.1 Methods .....	50
§3.3.2 Results.....	51
<b>CHAPTER 4. LEXICAL VARIATION, COGNITIVE REPRESENTATION, AND SPEAKER MOTIVATIONS.....</b>	<b>55</b>
§4.1 Behavioral data and facets of cognitive representation .....	56
§4.2 Methods .....	59
§4.3 Predictions based in an accumulative model.....	60
§4.4 Predictions based in competing motivations .....	62
<b>CHAPTER 5. SPEAKER MOTIVATIONS AND LEXICAL INNOVATION.....</b>	<b>66</b>
§5.1 Methods .....	67
§5.2 Predictions .....	71
§5.3 Model of innovative utility .....	73
<b>CHAPTER 6. INNOVATIVE UTILITY HISTORICALLY.....</b>	<b>80</b>
§6.1 Consequences of propagation in the speech community.....	81
§6.1.1 Methods .....	81
§6.1.2 Results.....	84
§6.2 Novel semasiological and innovative utility.....	85

§6.2.1 Methods .....	86
§6.2.2 Results.....	89
<b>CHAPTER 7. CONCLUSION.....</b>	<b>92</b>
§7.1 Degree of polysemy historically.....	92
§7.2 Speaker motivations and lexical development historically .....	93
§7.3 The effects of novelty historically .....	95
§7.4 Methodological implications .....	96
§7.4.1 Environmental distribution.....	96
§7.4.2 Synchrony, diachrony, and history .....	98
§7.4.3 Limitations.....	99
§7.5 Discussion.....	100
<b>REFERENCES.....</b>	<b>101</b>



## LIST OF FIGURES

Figure 2.1: Discrete probability distributions of degree of polysemy in the common lexicon historically.....	27
Figure 2.2: Discrete probability distributions of degree of polysemy in the core lexicon historically .....	30
Figure 2.3: Discrete probability distributions of degrees of polysemy in the core lexicon historically ranging from 2 to 5.....	31
Figure 3.1: Partial effects of the significant predictors in the multiple regression model fitted to degree of polysemy .....	53
Figure 4.1: Correlations between response times in lexical decision and features of lexical variation .....	61
Figure 4.2: Factorial structure among features of lexical variation .....	63
Figure 5.1: A model of innovation in comparison to a model of polysemy.....	69
Figure 5.2: Partial effects of the significant predictors in the multiple regression model fitted to lexical development historically .....	76
Figure 5.3: Partial effects by frequency bin.....	79
Figure 6.1: Schematic for a diachronic relationship between processes of lexical development and changes in distributional variation .....	82
Figure 6.2: Innovative utility relative to semasiological structure by generation .....	88
Figure 6.3: Probability density plots for changes in innovative utility from $t$ to $t+1$ by change type .....	90

## LIST OF TABLES

Table 2.1: Five example word forms/entries extracted from three dictionaries .....	24
Table 2.2: Degrees of polysemy historically for five example word forms .....	25
Table 2.3: Average degree of polysemy in common lexicon by publication date .....	27
Table 2.4: Average degree of polysemy in core lexicon by publication date .....	29
Table 3.1: COHA composition by constructed generation .....	34
Table 3.2: Frequency, dispersion, and dissemination .....	38
Table 3.3: Relative frequency distribution across genre for the corpus at $t_7$ .....	40
Table 3.4: Relative frequency distributions of <i>chaos_n</i> and <i>hypothesis_n</i> across genre at $t_7$ .....	41
Table 3.5: Ten most frequent local environments for <i>anxiety_n</i> at $t_7$ .....	43
Table 3.6: $w \times E$ and $q \times E$ for the ten most frequent local environments of <i>anxiety_n</i> at $t_7$ .....	45
Table 3.7: Core lexicon by part of speech.....	46
Table 3.8: Correlations (Pearson's $r$ ) among features of distributional variation historically .....	48
Table 3.9: Collinearity ( $\kappa$ ) by generation .....	49
Table 3.10: Word forms by generation.....	50
Table 3.11: Summary of model of polysemy .....	52
Table 5.1: Polysemy and lexical development historically .....	68
Table 5.2: Predicted characteristics of forms used in innovation per factor structure and theory .....	72
Table 5.3: Distribution of change type by generation .....	74

Table 5.4: Summary of model of lexical innovation.....	75
Table 6.1: Summary of model of propagation.....	84
Table 6.2: Innovative utility scores historically.....	87
Table 7.1: Features of distribution by model.....	97

## §1. INTRODUCTION.

### §1.1 Identification of the problem.

This dissertation investigates the role of speaker motivations in processes of innovation as reflected in the types of forms speakers recruit to create novel form-meaning pairs. As Bybee (2015:195) notes, “There is an interesting tension between the need for words to be stable in their meaning so that language users understand each other and the tendency and need to adapt old words to new uses.” From this perspective, this dissertation considers the usage-based (ie, corpus-based) characteristics of the “old words” that speakers re-purpose to “new uses.” Evidence of this particular proclivity among speakers historically is documented in any dictionary; Britton (1978) estimates that roughly half (44%) of forms in the English mental lexicon are polysemous to some degree.

Example (1) presents the full dictionary entry for the form *bedfellow\_n*<sup>1</sup>, taken from the 1913 edition of Webster’s Dictionary (Porter 1913). The form *bedfellow\_n* as described in 1913, then, is monosemous; in other words, it is only used to express a single concept.

- (1) *bedfellow\_n*      1913      One who lies with another in the same bed

---

<sup>1</sup> The nomenclature used throughout this study when referring to word forms is *lemma\_part of speech* in lowercase.

Example (2), on the other hand, presents the full dictionary entry for *bedfellow\_n* as described in the 2014 edition of the Merriam-Webster's (*Merriam-Webster*) online dictionary; as described more or less contemporarily, *bedfellow\_n* denotes two concepts.

- |     |                    |      |                                   |
|-----|--------------------|------|-----------------------------------|
| (2) | <i>bedfellow_n</i> | 2014 | One who shares a bed with another |
|     |                    | 2014 | Associate; ally                   |

From this perspective, we define “lexical innovation” as the development of novel form-meaning pairs between two points in time; this dissertation considers speaker motivations underlying the creation of such novelty.

### **§1.2 Processes underlying lexical innovation.**

The novel form-meaning pair associated with *bedfellow\_n* in (2) is ultimately a product of two distinct processes: the creation of the novel form-meaning pair and the propagation of that novelty in the speech community (Croft 2000, Nerlich & Clarke 1989). At some point in the last century, some speaker (or speakers) introduced a novel way to express the concept “ally/associate” using the extant form *bedfellow\_n*; over time, additional speakers began adopting *bedfellow\_n* as a way to express the concept “ally/associate.” At some point, enough speakers adopted this particular usage that it became a conventional (ie, non-novel and familiar) way to express the concept

“ally/associate” in the speech community as attested by its inclusion in a dictionary.

From this perspective, the act of innovation is a synchronic one, while the propagation of that novelty develops diachronically.

The cognitive processes that facilitate lexical innovation of this variety are predominantly metaphor and metonymy (Nerlich & Clarke 1989, Geeraerts 2015).

Metaphor is an analogical process in which (some element of) of one concept is construed in terms of (some element of) another concept (Traugott & Dasher 2005). The novel meaning associated with *bedfellow\_n* in (2) is sourced in metaphor; in this case, sharing a common agenda is construed in terms of sharing a bed. In the parlance of metaphor theory, sharing a bed is considered the source concept while sharing an agenda is the target concept.

Metonymy is a similar process in which the construed relatedness between source concept and target concept is sourced in “conceptual contiguity” (Blank 2003). An example of the metonymic process at work is presented in (3) and (4). Example (3) presents the only entry associated with the form *barbecue\_n* in Webster’s 1828 edition; example (4) presents an added entry from Webster’s 1913 edition. In this case, the “hog roasted whole” (ie, meat) is construed as the “social entertainment” surrounding the eating of meat.

- (3) *barbecue\_n* 1828 A hog roasted whole
- (4) *barbecue\_n* 1913 A social entertainment, where many people assemble, usually in the open air, at which one or more large animals are roasted or broiled whole

Such processes are often necessitated by some type of (construed) communicative deficiency in the lexicon, including gaps in the mental lexicon (ie, conceptual categories without referring forms), the emergence of new conceptual categories via socio-cultural change, and the need for a higher degree of specification within existing conceptual categories (Nerlich & Clarke 1989, Blank 2003).

### **§1.3 Speaker motivations.**

The causal mechanisms underlying these processes of innovation relate to the speaker's communicative goals, or motivations, at the time of innovation. While traditionally these motivations have been dichotomized as economy and expressiveness, Croft (2000) motivates a three-way distinction that additionally includes "not being misunderstood." From this perspective, speakers theoretically create novel form-meaning pairs with the intention of being economical, being expressive, and/or not being misunderstood.

Two decidedly different accounts have been presented in the literature regarding the relative prominence of these intention types in processes of innovation. Accounts based directly in the conceptual workings of metaphor and metonymy frame innovation as based in expressiveness and not being misunderstood; usage-based accounts, however, based in relationships between frequency and lexical access, on one hand, and frequency and degree of polysemy, on the other hand, frame innovation as predominantly economically motivated.

We consider the latter account first to introduce the types of features of (corpus-based and behavioral) lexical variation that have been considered in the literature in relation to speaker intention, and that will be considered throughout this dissertation. These two accounts ultimately make very different predictions regarding the lexical characteristics of forms used in innovation processes.

### **§1.3.1 Economy.**

Economy-based accounts of speaker motivation in processes of innovation generally take as their starting point the synchronic relationship between between corpus frequency and degree of polysemy. Zipf (1945) demonstrates a strong positive association between frequency and degree of polysemy in the mental lexicon; such findings have been corroborated using more balanced corpora, as well as more recent dictionary resources (eg, Paivio et al. 1968, Hay 2001, Hoffman et al. 2013). Linguists



have often interpreted this synchronic relationship as evidence of the role of frequency in processes of lexical innovation historically.

The relationship between frequency and economic motivations, then, is ultimately based in the relationship between frequency and lexical access. Balota & Chumbley (1984) demonstrate a strong negative association between frequency and lexical decision times, which persists across large-scale datasets of lexical decision data (Balota et al. 2007). This findings has generally been interpreted as evidence for role of frequency in the cognitive process of entrenchment, a process in which the mental representation of a linguistic form is strengthened via lexical access. An implication (or result) of this process is that highly entrenched forms (ie, more strongly represented forms) are more likely to be repeated by speakers (Bybee 1985). As noted in Schmid (2014:10), “this gives rise to a feedback loop in which frequency comes to serve as both a cause and effect of entrenchment.”

The influence of this type of “feedback loop” on processes of lexical development amounts to the following: if strongly represented words are more likely to be used by speakers, then they are more likely to be used when speakers innovate. An example of this type of reasoning is presented in Lee (1990:212), who cites that more frequent forms have “greater opportunity to be applied to novel domains,” and hence are more likely to develop novel meaning. Hay (2003:107) takes a more definitive stance, stating that “once a form is sufficiently frequent, then meaning proliferates.”

The effects of degree of polysemy are not limited to high lexical frequency. Degree of polysemy is also reflected synchronically in how lexica distribute across

various linguistic environments that comprise a corpus. Koehler (1986), for example, demonstrates that more polysemous forms tend to be found in a higher proportion of corpus texts. Similarly, if we consider the local environments in which lexica distribute (ie, some NxN window of surrounding text in a corpus), more polysemous forms tend to distribute more evenly across these environments (McDonald & Shillcock 2001); they also tend to distribute in local environments that are qualitatively more varied (Sagi et al. 2011, Hoffman et al. 2013). Lastly, Zipf (1945) observes that highly polysemous forms exhibit a tendency to be shorter orthographically.

To account for the historical development of this set on synchronic relationships, Fenk-Oczlon & Fenk (2010:105) propose an “interactive step-up” between frequency and polysemy akin to the “feedback loop” between frequency and strength of representation. The authors argue that “frequent use favours the tendency to shortness and polysemy, and shortness and polysemy favors frequent use for obvious reasons — the use of shorter expressions is economically motivated, and words encoding a higher number of meanings fit in a higher number of contexts” (Fenk-Oczlan & Fenk 2010:105).

Hay (2003:57) adds a dimension to (or perhaps an implication of) this “step-up,” stating that “If every meaning of a word has some possibility of spawning new meanings, then the more meanings a word has, the more likely it is to acquire still further meanings. As new meanings are acquired, the possibilities for further meanings grows exponentially.”

As to whether these two “step-ups” are individually motivated, or if one feeds the other, is unclear; in either case, taken collectively they predict that when speakers create

novel form-meaning pairs, they recruit forms that are cognitively “on-hand” and forms with high degrees of polysemy. We will refer to this collection of predictions based exclusively in the speaker economy as an “accumulative” model of lexical innovation.

### **§1.3.2 Expressiveness and not being misunderstood.**

Economy-based accounts of speaker motivation, then, are largely based on the usage-based characteristics of highly polysemous word forms synchronically. In contrast, the motivations of expressiveness and not being misunderstood are based in the conceptual underpinnings of the processes that facilitate innovation: metaphor and metonymy.

Hopper & Traugott (1993:65) define metaphor and metonymy as processes motivated by the speaker’s need to be expressive. From this perspective, they are generally viewed as processes by which conceptual categories are enriched semantically. Nerlich & Clarke (1989:128) refer to such enrichment as “additional semantic power.” The novel form-meaning pairs associated with *bedfellow\_n* in (2) and *barbecue\_n* in (4) provide examples of this enrichment process within the conceptual categories of “ally/associate” and “picnic/outdoor party,” respectively.

Theoretical constraints on processes of innovation would also seem to motivate the recruitment of more informative and expressive forms. Lakoff & Johnson (1980) theorize that source concepts will be conceptually more concrete than target concepts; in this way, processes of innovation are viewed as processes in which “concrete concepts

are employed in order to understand, explain, or describe less concrete phenomena” (Heine et al. 1991:28).

While “concreteness” in theory of metaphor research is a relative notion, Katz (1989) investigates the influence of subjective concreteness ratings on the types of forms speakers recruit in innovation processes. In a forced decision task, the author asked subjects to complete “metaphorical expressions” using a word from a predetermined list (eg, “the USA is the [*blank*] of nations”). The author found that speakers generally recruit more concrete forms in completing this particular innovative task.

How cognitive variables (or constructs) like informativity and expressivity manifest in a corpus in terms of features of distributional variation, on the other hand, is less clear from the literature. Fenk-Oczlon & Fenk (2010) hypothesize that informativity may be sourced in local co-occurrence, specifically in concentrated local environments. Fenk-Oczlon & Fenk (2010) provide the example of potentially frequent adjectival predications (eg, “The fox is cunning”), and argue that speakers may exploit such frequent patterns in local co-occurrence in processes of lexical innovation (eg, “John is a fox”). While perhaps not the best example, this particular interpretation frames lack of variation (ie, consistency) in the types of environments speakers experience lexica as a potential source of informativity.

Lastly, arguments exist in the literature that suggest high degrees of polysemy may detract from the informativity of forms in the lexicon. Nerlich & Clarke (1989) argue that it is ultimately concepts expressed by highly polysemous forms that motivate speakers to innovate in the first place. From this perspective, while a novel

form-meaning pair brings added informativity to a conceptual category, additional semasiological structure theoretically detract from the collective informativity of the lexical unit (Polikarpov 2006). While largely speculative, this line of argumentation would seem to make intuitive sense; although not explicitly proposed in the literature, implicit in this particular characterization is that forms with lower degrees of polysemy are more informative.

While less prominently than expressiveness, the role of the speaker intention of not being misunderstood has also been proposed in the literature as relevant to processes of metaphor and metonymy. Fenk-Oczlon & Fenk (2010) argue that that speakers recruit cognitively familiar forms to ensure the understandability of innovative form-meaning pairs, and avoid less familiar forms. They propose that familiarity as a construct is sourced distributionally in frequency.

### **§1.3.3 Competing motivations.**

The purported features of lexical variation relevant to speaker intention in the process of innovation across intention types, then, are quite different, and largely at odds. Features relevant to the speaker's need to be economical are described cognitively in terms of accessibility; such forms are highly frequent, highly polysemous, and shorter in word length; they also distribute across a broader range of linguistic environments. Features relevant to the speaker's need to be expressive in processes of innovation are

described cognitively in terms of informativity; such forms are characterized as less polysemous and more concrete subjectively; it has also been suggested that informative forms distribute in more concentrated local environments. Lastly, features relevant to the speaker's need to not be misunderstood are described cognitively as familiar; familiar forms are characterized as highly frequent. In this way, frequency as a feature of lexical variation seems relevant to both the intention of being economical and not being misunderstood.

Theoretically, the economy-based model presented in §1.3.1 is without constraints; growth begets growth, as growth strengthens representation. Implicit in this model is that any feature of lexical variation that facilitates lexical access will facilitate innovation; shorter word length (New et al. 2006), higher degrees of polysemy (Rodd et al. 2002), more distributed local environments (McDonald & Shillcock 2001), and more dispersed text environments (Adelman et al. 2006) have all been demonstrated to facilitate lexical access, independently of frequency. From this perspective, there remains very little room for alternative interpretations of how different features of lexical variation may relate to competing motivations.

While accounts based in economy and accounts based in not being misunderstood and expressiveness seem largely incompatible, however, they likely just reflect different research interests. Linguists interested in metaphor focus on the role of expressiveness and the types of communicative needs arising in conceptual categories; linguists interested in speaker economy focus on accessibility and the types of forms that are cognitively on-hand. As a result, economy-based accounts largely neglect the

motivations underlying metaphor and metonymy, while expressiveness-based accounts largely neglect the realities of repeated usage and its effects on speaker and cognitive representation. A unified account of potential competition among these three intention types has not been presented in the literature.

#### **§1.4 Speaker and speech community methodologically.**

Both accounts of how speaker intentions may influence the lexical characteristics of forms speakers recruit in innovation processes are largely speculative (with the exception of Katz (1989)). While “interactive step-up” accounts based in synchronic relationships make empirically-testable predictions, the relationships themselves do not provide evidence of the role of economy (or any other intention for that matter) in lexical development which has happened historically for (at least) two reasons:

- (1) Degree of polysemy and the development of novel semasiological structure are distinct phenomena: a novel form-meaning pair develops between two points in time,  $t$  and  $t+1$ ; degree of polysemy is the aggregate of these developments that preceded any given  $t$  (ie, historically with respect to  $t$ );
- (2) The lexical variation relevant to cognitive processes of innovation is the lexical variation at the time of innovation,  $t$ , not the lexical variation after the innovation

has become an established form-meaning pair in the speech community at  $t+1$  (Haspelmath et al. 2014).

In other words, synchronic speakers (and their representations of forms in the lexicon as approximated by a synchronic corpus) have nothing to do with extant semasiological structure in the mental lexicon that has developed historically. Theoretically, speakers at  $t$  are only responsible for changes that occur from  $t$  to  $t+1$ ; more specifically, speakers innovate at  $t$  and novelty develops in the speech community from  $t$  to  $t+1$  (when propagation is successful).

From a methodological perspective, then, accounts of speaker intention based in synchronic relationships confound the synchronic speaker (and cognitive representation) with the diachronic effects of successful propagation in the speech community (multiple times over). A more realistic corpus-based approximation of the types of forms speakers exploit in innovative processes requires an independent treatment of speaker and speech community; it also requires the ability to approximate lexical development between  $t$  and  $t+1$  independently from degree of polysemy at  $t$ . The methodological argument presented here is that our understanding of the motivations underlying innovation can be better informed by accounting for these distinctions in our methods.



## **§1.5 Dissertation organization.**

This dissertation is organized as follows:

Chapter 2 introduces a dictionary-based approach to operationalizing the degree of polysemy at multiple synchronic stages historically. Comparisons of probability distributions for degree of polysemy across the three dictionaries demonstrate, first, that lexicographic practices have remained rather consistent historically, and second, that substantial semasiological development has occurred over the past two centuries in American English.

Chapter 3 investigates the synchronic consequences of degree of polysemy in terms of distribution in a speech community historically. Five features of distributional variation are described, including frequency, word length in syllables, and three dimensions of environmental distribution — text environments (ie, dissemination), discourse environments (ie, genre), and local environments. Results from a multiple linear regression model demonstrate that higher degrees of polysemy in the mental lexicon are reliably reflected synchronically in higher degrees of frequency, shorter word length, and more variability across discourse and local environments; importantly, these relationships persist historically.

Shifting focus, chapter 4 considers the features of distributional variation from the perspective of the speaker and cognitive representation. The relationships among these features and three features of behavioral variation — response times in lexical decision, concreteness ratings, and age-of-acquisition ratings — are investigated to better

understand the types of information captured by distributional features, as well as to better understand the distributional manifestations of the cognitive variables (ie, accessibility, familiarity, and informativity) underlying speaker intentions (ie, economy, not being misunderstood, and expressiveness) in processes of lexical innovation. Factor analysis of these nine features of lexical variation identifies three orthogonal factors that loosely align with these cognitive constructs; such characterizations serve as the basis for the predicted relationship between distributional variation and speaker intention in a model based in competing motivations.

Chapter 5, then, introduces a temporally idealized model of lexical development, in which the speaker exists synchronically at  $t$  and the speech community exists diachronically from  $t$  to  $t+1$ . The role of speaker motivations in processes of innovation is investigated by considering how cognitive representation (as approximated distributionally) at  $t$  influences the likelihood of development of novel semsiological structure from  $t$  to  $t+1$  (as approximated by a comparative dictionary methodology). Results from a logistic regression model demonstrate evidence of competition among speaker intentions historically, suggesting roles for all three motivations in processes of innovation. Results are importantly distinct from predictions based in both a model of polysemy and the theoretical accumulative model.

Chapter 6 considers the aggregate distributional effects of the propagation of novel form-meaning pairs from  $t$  to  $t+1$ , first, from the perspective of the speech community, and second, from the perspective of the speaker and cognitive representation. In the case of the former, results from a panel analysis demonstrate that the development

of novel semasiological structure is accompanied by increases in frequency, increases in degree of dissemination, and increases in discourse environments; surprisingly, such changes are not reflected in novel local environments.

In the case of the latter, comparison of innovative utility (ie, likelihood) scores derived from the logistic regression model from chapter 5 demonstrate that forms that develop novel form-meaning pairs from  $t$  to  $t+1$  have lower innovative utility at  $t+1$  than at  $t$ . While forms are theoretically more accessible at  $t+1$  via increases in frequency (and degree of polysemy), these gains are presumably outweighed by losses in informativity sourced environmentally, making forms less amenable for innovation at  $t+1$ . Results ultimately account for why an accumulative model does not hold historically, and provide additional evidence of the effects of competing motivations historically.

## **§2. POLYSEMY HISTORICALLY.**

In order to begin to address the research goals of this dissertation, we need to be able to approximate the degree of polysemy in the mental lexicon historically. The first section of this chapter considers two different approaches to approximating degree of polysemy: a corpus-based approach and a corpus-external approach. An argument is presented against corpus-internal approaches and for the utility of a corpus-external approach. The second section describes a dictionary-based methodology to approximate degree of polysemy historically. The third section investigates the comparability of the dictionaries in terms of consistency in lexicographic practices. Lastly, the fourth section considers aggregate semasiological development in the mental lexicon over the past ~185 years.

### **§ 2.1 Methodological approaches to degree of polysemy and lexical change.**

#### **§2.1.1 Corpus-based approaches.**

Corpus-based approximations of degree of polysemy generally take the form of a vector space model (VSM). Traditional VSMs are based theoretically in the “distributional hypothesis,” which states that words with similar meanings distribute contextually in similar ways (Harris 1954). Such models are generally implemented via

the construction of a matrix, dubbed here as a “word-environment” matrix (although see Turney & Patel (2010) for a more detailed taxonomy of VSMS). Rows in the matrix correspond to words; columns in the matrix correspond to some type of linguistic environment, generally a document or a window of text surrounding some target word (Lund & Burgess 1996).

If we consider a corpus as comprised of a set of linguistic environments, a VSM can be used to represent the overall frequency of any word,  $w$ , in terms of its relative frequencies across each linguistic environment,  $E$ , in the corpus. We can then compare word vectors  $w_1 \times E$  and  $w_2 \times E$  to assess semantic similarity between  $w_1$  and  $w_2$ . Rapp (2003) provides support for the utility of VSMS in capturing semantic similarity: the author demonstrates that VSMS outperform the average human test-taker in the multiple-choice synonym portion of the TOEFL exam.

Traditional applications, then, are onomasiological in nature, and only useful in comparing  $w_1$  to  $w_2$ . However, more recent applications of VSMS (eg, Sagi et al. 2011, Hoffman et al. 2013) have been developed based in a variation of the distributional hypothesis, dubbed here as the “semasiological distributional hypothesis.” This hypothesis states that greater distributional variation in a corpus implies greater semasiological range, and is based in the conceptualization of a form-meaning pair as a “context of usage.” Instead of comparing word vectors  $w_1 \times E$  and  $w_2 \times E$  to assess semantic similarity between  $w_1$  and  $w_2$ , semasiological approaches, then, attempt to quantify variation (or dissimilarity) of the linguistic environments,  $E$ , in which a given  $w$

occurs. Such variation is assumed to reflect multiple contexts of usage, and hence, multiple meanings.

The problem with such approaches, however, is the extent to which a linguistic environment aligns with the linguistic notion of a “context of usage.” On one hand, while it is reasonably clear what these models capture in terms of the math, it is less clear how the math is best interpreted from the perspective of meaning. On the other hand, a “context of usage” is decidedly more than just local co-occurrence; presumably of equal importance are discourse-pragmatic features not capturable in a corpus.

These problems become more apparent when considered from the perspective of lexical development historically. Gulordava & Baroni (2011), for example, demonstrate that patterns of local co-occurrence can change without being symptomatic of any change semantically; as the authors note, changes in local environments can additionally reflect changes in the overall composition of the lexicon or changes in content of the socio-cultural conversation. Gulordava & Baroni (2011:70) refer to such changes as “false-positives.” Similarly, change in linguistic environments can reflect internal changes in semasiological structure, and not necessarily new semasiological structure.

### **§2.1.2 Corpus-external approaches.**

The traditional corpus-external approach to approximating degree of polysemy is based in counts of dictionary definitions; from a synchronic perspective, this approach is

well-established in the literature. Psycholinguistic research generally follows such an operationalization when investigating the role of semasiological structure in lexical decision and naming tasks (eg, Baayen et al. 2006). WordNet synsets have become the norm in this regard, but alternative, more traditional dictionary resources (eg, Webster's 1913) continue to be used as well (Hay 2003).

Corpus-external approaches to degree of polysemy, however, are not without criticisms. Some linguists, for example, have questioned the cognitive validity of the notion of a word "sense" as traditionally conceptualized in dictionaries, and the extent to which word senses can realistically be delineated and enumerated (Kintsch 2007, Kilgarriff 1997). Kilgarriff (1993:366) additionally cites the challenges of aligning a token of meaning in a corpus to "one and only one" meaning in a dictionary, even in cases where corpus meaning is straightforward.

However, as noted by Hay (2003:53), we do not necessarily have to assume "a one to one relationship between mental representations and the content of dictionary entries" (Hay 2003:53). Instead, the number of definitions associated with a given *w* more likely reflect what Hanks (2000) and Geeraerts (2014) refer to as "meaning potential", or "the possibility to express a flexibly defined range of meanings when they are put to use in a given context" (Geeraerts 2015:417).

### **§ 2.1.3 Arguments for a corpus-external approach.**

Based on the challenges associated with validating the semasiological distributional hypothesis, a corpus-external approach to approximating degree of polysemy historically is implemented here. While several (corpus-based) environment-based metrics are considered in this study, we treat such metrics as distinct from degree of polysemy and make no assumptions regarding the semasiological distributional hypothesis.

The argument presented here is that an operationalization based in lexicographic description is ultimately more useful with respect to the tasks at hand, specifically the identification of novel form-meaning pairs, than vector space models. While corpus-external approaches have not been applied to the issue of lexical development at scale, such a methodology is presented in the next section.

## **§2.2 An historical dictionary methodology.**

This dissertation utilizes three dictionary resources to describe the degree of polysemy in the mental lexicon at three points in time in the history of American English. These three dictionaries include: Webster's 1828 Edition (Webster 1828), Webster's 1913 Edition (Porter 1913), and Merriam-Webster's present-day online dictionary (Merriam-Webster 2014). The latter is based on the print version of the 11th edition of



the Merriam-Webster's Collegiate Dictionary. Both Webster's 1828 and 1913 editions are available as text files via Project Gutenberg. These particular resources were selected for three reasons: (1) their availability in machine-readable formats, (2) their common source of publication, and (3) the similarity of their respective intended audiences, assumed here to be more or less a "general" one.

A lexical unit (or word form) is defined throughout this dissertation in terms of the lemma and part-of-speech. Lexical units and their respective senses were extracted from the three dictionaries in two different ways. In the case of Webster's 1828 and 1913, these data were simply extracted from the publicly available text files. For 1828 and 1913, then, we have full access to dictionary content. In the case of Merriam-Webster's online dictionary, data were extracted via web scraping methods. A simple script was written to automate search and entry extraction from the online dictionary.

As an online, search-based resource, however, we only have access to entries that are explicitly searched for; for our purposes here, this search was limited to lexical units occurring at greater than 0.25 parts per million (ppm) in the Corpus of Historical American English (COHA) (Davies 2010-) for the three decades in which the dictionaries were published (ie, 1820s, 1910s, and 2000s). The frequency threshold of 0.25 ppm is largely arbitrary; however, it serves to narrow the domain of inquiry to forms commonly used by speakers. This particular subset of the lexicon will be referred to as the "common" lexicon.

In all three dictionaries, distinct form-meaning pairs associated with a given form are explicitly numbered; this numbering system was exploited across all three resources to delineate and subsequently count the number of senses associated with each form in the common lexicon. Before aggregating over form-meaning pairs to get at actual counts, however, several sense types were excluded; these include form-meaning pairs identified as “obsolete” or “obscure”, following Hay (2003:53). Form-meaning pairs identified/labeled as “regional” were additionally excluded; the reason for doing so is that in most cases this particular label referred to form-meaning pairs predominant in English-speaking countries outside of the United States. In the case of the 2014 dictionary, only senses included in the main entry were included; sub-entries labeled as “Medical” and “Biographical”, for example, were excluded.

The degree of polysemy for approximately 20 thousand forms in the common lexicon can be described (ie, are included) in all three dictionaries; this subset of the common lexicon will be referred to as the “core” lexicon. Five example lexical units from the core lexicon and their associated form-meaning pairs from each dictionary are presented in Table 2.1.

*Table 2.1: Five example word forms/entries extracted from three dictionaries.*

Form	Year	Entry	Definition
mundane_a	1828	1	Belonging to the world; as mundane sphere; mundane space
	1913	1	Of or pertaining to the world; worldly; earthly; terrestrial
	2014	1	Of, relating to, or characteristic of the world
	2014	2	Characterized by the practical, transitory, and ordinary; commonplace
handcuff_v	1828	1	To manacle; to confine the hands with handcuffs
	1913	1	To apply handcuffs to; to manacle
	2014	1	To apply handcuffs; to manacle
	2014	2	To hold in check; make ineffective or powerless
disappointment_n	1828	1	Defeat or failure of expectation, hope, wish, desire or intention; miscarriage of design or plan
	1913	1	The act of disappointing, or the state of being disappointed; defeat or failure of expectation or hope; miscarriage of design or plan; frustration
	1913	2	That which disappoints
	2014	1	The act or an instance of disappointing; the state or emotion of being disappointed
	2014	2	One that disappoints
anteroom_n	1828	1	A room before or in front of another
	1913	1	A room before, or forming an entrance to, another; a waiting room
	2014	1	An outer room that leads to another room and that is often used as a waiting room
adulterate_v	1828	1	To commit adultery
	1828	2	To corrupt, debase, or make impure by an admixture of baser materials
	1913	1	To corrupt, debase, or make impure by an admixture of a foreign or baser substance
	2014	1	To corrupt, debase, or make impure by the addition of a foreign or inferior substance or element

### §2.3 Evidence of lexicographic consistency historically.

Table 2.2 summarizes Table 2.1 in terms of counts of form-meaning pairs, or degrees of polysemy, for the five example lexical units. Ideally, we want to be able to use the synchronic descriptions presented in Table 2.2 to evaluate change diachronically. While the consistency in language across the three resources attested in Table 2.1 provides some evidence that lexicographic practices have remained consistent historically, such an “eyeball” test does not reasonably scale to a full lexicon.

*Table 2.2: Degrees of polysemy historically for five example word forms.*

Form	1828	1913	2014
mundane_a	1	1	2
handcuff_v	1	1	2
diappointment_n	1	2	2
anteroom_n	1	1	1
adulterate_v	2	1	1

To approach the issue of lexicographic consistency more systematically, we compare probability distributions for degrees of polysemy in the common lexicon at each decade of dictionary publication (ie, 1820s, 1910s, and 2000s). A random sample of 1,000 lexical units from the common lexicon was independently generated for each decade;

sample independence serves to control for potential variation across the three dictionaries attributable to lexical change.

For each sample, dictionary searches were performed for all forms. If a form was included in the dictionary, the number of associated definitions was assigned as the degree of polysemy for the form. If a form was not included in the dictionary, the form was assigned a degree of polysemy equal to monosemy; Hay (2003) follows a similar procedure. For the 2000s sample, the online search and extraction method described in §2.2 was employed.

The null hypothesis is that the distributions of these three samples come from some common “parent” distribution; in other words, the null hypothesis is that the three distributions do not differ. If there has been no change in lexicographic practices historically, then we would expect the null hypothesis to hold. Zipf (1949) notes that degree of polysemy distributions have properties akin to power law distributions; informally, this means that most words in the lexicon have low degrees of polysemy, while a small number of words have high degrees of polysemy. For our purposes here, however, it ultimately does not matter whether attested distributions have power law characteristics or not; our only expectation is that they look the same.

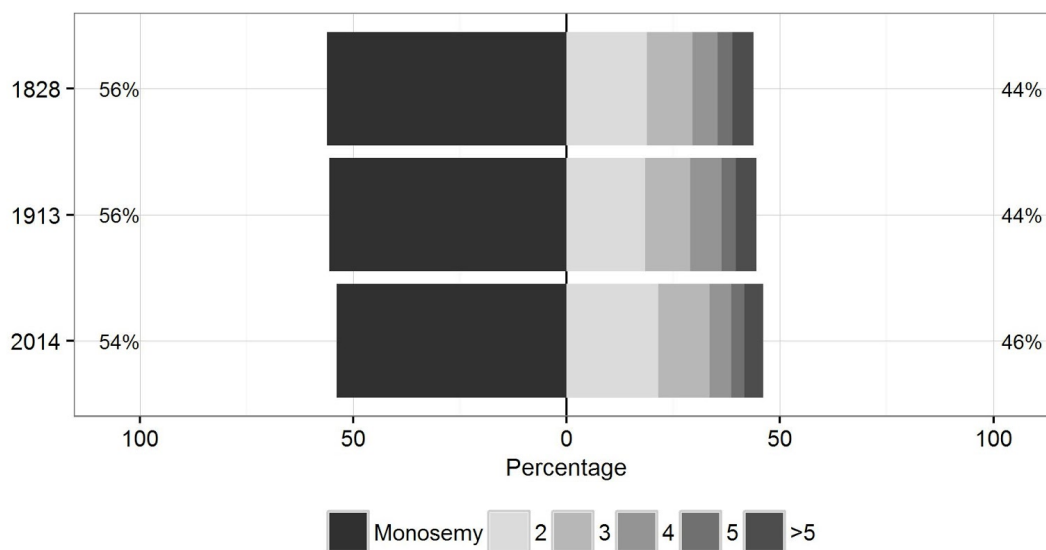
Table 2.3 summarizes the average degree of polysemy for the random samples by publication date. Paired Wilcoxon signed rank tests comparing temporally contiguous dictionary distributions (ie, 1828-1913 and 1913-2014) suggest that these means do not differ across dictionaries ( $p > 0.05$ , for both comparisons).

*Table 2.3: Average degree of polysemy in common lexicon by publication date.*

1828	1913	2014
2.29	2.29	2.15

Figure 2.1 summarizes the probability distributions for the three random samples. The portion of the stacked probability barplot to the left of origin illustrates the proportion of each lexicon comprised of monosemous forms. To the right of origin illustrates the portion of each lexicon comprised of forms with degrees of polysemy ranging from 2 through 5 and greater than 5. While there is some variation, the three distributions appear remarkably similar.

*Figure 2.1: Discrete probability distributions of degree of polysemy in the common lexicon historically.*



The Anderson-Darling k-sample test is used to test the hypothesis that a set of samples come from a common but unspecified distribution function (Scholz & Stephens 1987). The application of this test to the three distributions demonstrates no evidence to reject the null hypothesis ( $p > 0.05$ ), ie, degree of polysemy distributions do not differ across the three dictionaries. We interpret this result as evidence for consistency in lexicographic practices across the three centuries of lexical description, and as validation of a comparative dictionary methodology to the problem of lexical development historically.

#### **§2.4 Evidence of lexical development historically.**

Having demonstrated that lexicographic practice has remained rather consistent historically, in this section we consider three dependent samples (ie, the core lexicon) from the dictionaries to determine if lexical development has occurred historically. Degree of polysemy is observed for all forms in the core lexicon ( $n = 20,262$ ) in all three dictionaries; if development has occurred historically, we would expect the three distributions to be progressively more skewed towards higher degrees of polysemy from 1828 to 2014.

A simple comparison of the average degree of polysemy in the core lexicon across the three dictionaries provides preliminary evidence that the core lexicon has developed semasiologically. Table 2.4 below summarizes the average degree of

polysemy of the core lexicon by dictionary publication date. As can be noted, the core lexicon has become progressively more polysemous historically, with the average number of definitions per entry increasing from 2.47 in 1828 to 2.66 in 2014. Paired Wilcoxon signed rank tests comparing temporally contiguous dictionary distributions (ie, 1828-1913 and 1913-2014) suggest that these means do in fact differ across the centuries ( $p < 0.001$ , for both comparisons).

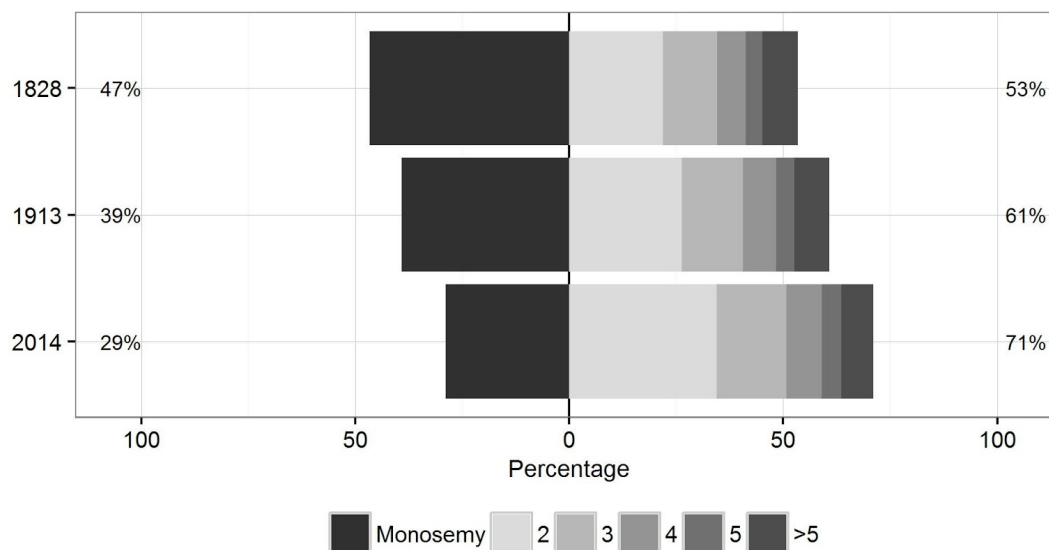
*Table 2.4: Average degree of polysemy in core lexicon by publication date.*

<b>1828</b>	<b>1913</b>	<b>2014</b>
2.47	2.55	2.66

Additional evidence of growth can be found when the discrete probability distributions of the core lexicon for each dictionary are compared. Figure 2.2 summarizes the probability distributions for the core lexicon over time; as the figure illustrates, the core lexicon has become progressively less monosemous in its constituency. While monosemous forms comprise approximately half (47%) of the core lexicon in 1828, such forms comprise only 29% of the core lexicon in 2014, some 185 years later.

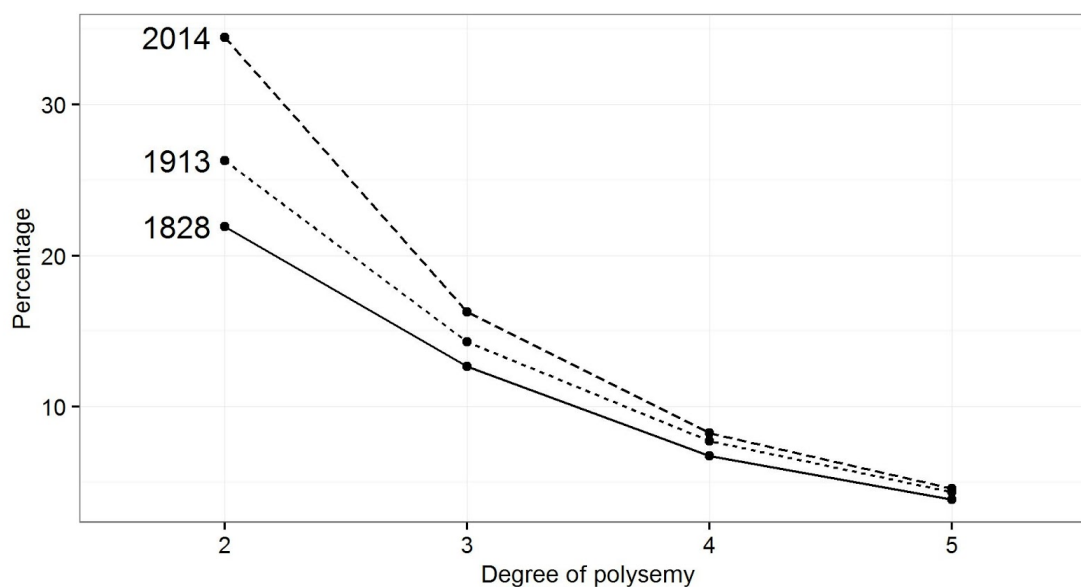


Figure 2.2: Discrete probability distributions of degree of polysemy in the core lexicon historically.



Results from an Anderson-Darling k-sample test suggest that these three distributions are not sourced in a common distribution function ( $p < 0.001$ ), ie, degree of polysemy distributions differ across the three dictionaries for the same set of forms. We interpret this result as evidence for semasiological development in the core lexicon historically. Figure 2.3 illustrates a more detailed perspective on the probability distributions of degree of polysemy for each dictionary, focusing on the distributions for degree of polysemy ranging from two to five. As can be noted, the differences in the composition of the core lexicon at 1828, 1913 and 2014 in terms of degree of polysemy speak intuitively to (non-random) development.

Figure 2.3: Discrete probability distributions of degrees of polysemy ranging from 2 to 5.



Taken collectively, different mean degrees of polysemy and differently sourced distribution functions provide strong evidence that speakers have been tasking out (or re-purposing) the core lexicon to create novel form-meaning pairs between 1828 and 2014. Per the findings presented §2.3, we can be fairly confident that the changes in the composition of core lexicon in terms of degree of polysemy are in fact a product of processes of lexical change, and not more detailed lexicographic practices historically. The remainder of this dissertation attempts to account for the development of this novelty historically.

### **§3. POLYSEMY AND THE SPEECH COMMUNITY.**

Having established a methodology for describing degree of polysemy in the mental lexicon historically, the goal of this chapter is to align these historical descriptions with historical speech communities. The first section of this chapter considers this alignment in terms of audience, specifically from temporal and rhetorical perspectives. The second section details the features of distributional variation considered in this study, as well as how these features align with different environments within (or facets of) the speech community. The third section presents an historical and synchronic model of polysemy, which demonstrates how polysemy is reflected within the speech community. While these characteristics have been well-described in the literature, a full model and an historical model are both lacking.

#### **§3.1 Aligning dictionaries and speech communities historically.**

The three dictionaries presented in §2 theoretically describe common and conventionalized form-meaning pairs used in the speech community at three distinct points in time. The goal of this section is to align these descriptions with actual usage in the speech community historically, as approximated by an historical corpus. In order to do this, we want to be fairly certain that the dictionaries and the texts comprising the

corpus share similar audiences, both rhetorically and temporally, such that we can be fairly certain the former is indeed a description of the latter.

From a rhetorical perspective, if the intended audience of the three dictionaries is assumed to be a general one, then ideally the corpus should be general as well, ie, comprised of texts from a variety of genres that best reflect everyday language use within a community of speakers. The Corpus of Historical American English (COHA) (Davies 2010-) is designed with this goal in mind, and would seem to align well with the sources of description. COHA is a 400 million word corpus comprised of over 100 thousand texts from 1810 to 2009. Constituent texts are drawn equally from newspapers, magazines and periodicals, non-fiction, and fiction. Alternative historical corpora, for example the Google n-gram corpus, are generally only comprised of books.

From a temporal perspective, if the intended audience of a given dictionary exists at some time,  $t$ , then ideally the texts comprising the corpus were generated during the same  $t$ . While the range of COHA (1810-2009) and the range of dictionary publication dates (1828-2014) are indeed comparable, observations in the two data sets occur at different intervals of time. Usage in COHA is observed at twenty decade intervals; degree of polysemy, however, is observed at only three (non-uniform) time intervals. Ultimately, the disparity between the two time-series makes audience alignment non-straightforward.

To best address this issue, twenty decades of COHA data are (re-) discretized as seven “generations.” Generation composition reflects an attempt to balance (1) audience similarity, (2) generation duration, and (3) generation size (in corpus tokens). The latter

becomes important when deriving count-based metrics per generation, especially for less frequent forms in the lexicon. Table 3.1 summarizes the composition of each generation.

The first row of Table 3.1 presents the number of the constructed generation. The second row summarizes each generation in terms of COHA decade composition. The third row provides the historic range of the generation. The fourth row provides the duration of the generation. The fifth row summarizes the size of the generation sub-corpus in total tokens. Finally, the sixth row illustrates how the three dictionaries map to the constructed generations.

*Table 3.1: COHA composition by constructed generation.*

$t_1$				$t_2$			$t_3$			$t_4$		$t_5$			$t_6$			$t_7$	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1810- 1849				1850- 1879			1880- 1909			1910- 1929		1930- 1959			1960- 1989			1990- 2009	
40 years				30 years			30 years			20 years		30 years			30 years			20 years	
36 mil				49 mil			59 mil			46 mil		70 mil			69 mil			55 mil	
W1828				--			--			W1913		--			--			W2014	

The three observations of degree of polysemy, then, align with three of the constructed generations in Table 3.1:  $t_1$ ,  $t_4$ , and  $t_7$ . As Table 3.1 illustrates, however, this alignment is imperfect: dictionary publication dates do not align precisely or uniformly with corpus-derived generations at  $t_1$ ,  $t_4$ , and  $t_7$ . The publication of Webster's 1828

(W1828), for example, occurs at the middle of  $t_1$ , while the publication of Webster's 1913 (W1913) occurs at the beginning of  $t_4$ ; the publication of Merriam-Webster's 2014 (W2014) occurs outside the historical range of  $t_7$ . Despite this lack of precision, however, the argument presented here is that publication dates and generation ranges are close enough to assume that the dictionaries at each  $t$  are a description of usage at each  $t$ . Per this design, degree of polysemy is not observed at generations  $t_2$ ,  $t_3$ ,  $t_5$ , and  $t_6$ .

Independent of dictionary publication dates, it should also be noted that the constructed generations do not occur at equally spaced intervals of time. Ultimately, this is a product of trying to balance generation duration and the constituency of each generation in terms of corpus tokens. Outside of  $t_1$ , generations are comprised of either twenty or thirty years of corpus data;  $t_1$  is comprised of forty years of data due to the sparsity of texts from the first half of the 19th century in COHA. As degree of polysemy is only observed at  $t_1$ ,  $t_4$ , and  $t_7$ , these generations will be the main focus of this dissertation.

### **§3.2 Distributional features of lexical variation.**

Again, we assume that COHA is representative of the speech community historically. The goal of this section is to present methods for describing how forms in the core lexicon distribute across the linguistic "environments" that comprise this speech community. Such a description will provide a lens from which to consider individual

relationships between usage within the speech community and degree of polysemy, processes of innovation, and the effects of novelty historically.

Per §1, several characterizations of usage within the speech community (ie, features of distributional variation) have been presented in the literature as relevant to these relationships; these include frequency, text dispersion, and metrics of local co-occurrence. Here, we consider frequency and a residual-based version of text dispersion, as well as a VSM-version of local co-occurrence. We also present a novel VSM-based approach to variation across genre. Lastly, we consider two features of lexical variation not based in distribution: part of speech and word length in syllables.

### **§3.2.1 Dissemination and frequency.**

We first consider environmental distribution from the perspective of the text. Corpora are comprised of texts; lexica vary in the extent to which they occur across such texts. As described in §2, VSM approaches to semantic similarity exploit frequency distributions across texts as proxies for word meaning. Usage-based approaches, on the other hand, use the text to approximate how dispersed a particular word,  $w$ , is within a community of speakers (eg, Altman et al. 2011). From this perspective, a text becomes a proxy for an individual language user; words found in more texts are viewed as being used by more speakers within the speech community, and hence more dispersed. Instead

of a distribution, then, “dispersion” as metric is a simple type count of texts containing a given  $w$  (Adelman et al. 2006).

As both Chesley & Baayen (2010) and Hoffman et al. (2013) note, however, corpus frequency and text counts are highly related; Hoffman et al. (2013) show that the two measures are correlated at (Pearson’s)  $r > 0.95$ . While excluding one of these variables from our characterization of distributional variation is an option, both variables are well entrenched within many usage-based linguistic research paradigms.

A fix to this problem is presented in Chesley & Baayen (2010): the authors regress frequency on text counts and use the residuals to represent frequency independent of text counts. For our purposes here, however, it makes more sense to regress text counts on frequency and use the residuals to represent text counts independent of frequency. In other words, the residuals become our proxy for text dispersion and frequency remains frequency. Frequency, then, is measured per 1 million tokens, or parts per million (ppm); these counts are counts of lemmas by part-of-speech, which include counts of all inflectional variants of a given  $w$ .

It is important to note that more current generations in COHA are comprised of more texts than earlier generations. This is a simple result of the fact that there are fewer texts available from the 19th century. As a result, it is possible that the relationship between frequency and dispersion is not uniform historically. For this reason, generation is controlled for when regressing (log) text counts on (log) frequency to calculate residuals.



We follow Adelman et al. (2011) in referring to this residual-based metric as “dissemination.” A dissemination value of greater than 0 means that a given  $w$  occurs in more texts than expected based on its level of frequency. A dissemination value of less than 0 means that a given  $w$  occurs in fewer texts than expected based on its level of frequency. Magnitude reflects the distance of the observed values from the fitted regression line.

Table 3.2 summarizes the frequency, dispersion, and dissemination values for two word forms at generation  $t_7$ : *environmental\_a* and *powerful\_a*. Note that while the frequencies of the two forms are essentially comparable, dispersion values are quite disparate: *environmental\_a* occurs in 4.9% of the 23,684 constituent texts at  $t_7$ , while *powerful\_a* occurs in 12.1%. Dissemination values (in log units) illustrate that both forms distribute textually in unexpected ways. A dissemination value of -0.226 suggests that *environmental\_a* occurs in fewer texts than expected, while a dissemination value of 0.603 suggests that *powerful\_a* occurs in more texts than expected.

*Table 3.2: Frequency, dispersion, and dissemination.*

Form	Frequency	Dispersion	Dissemination
<i>environmental_a</i>	68.4 ppm	4.9%	-0.226
<i>powerful_a</i>	74.5 ppm	12.1%	0.603

Forms with high dissemination values, then, are interpreted as distributing across more text environments, while forms with lower dissemination values are interpreted as distributing across fewer (and less varied) text environments.

### §3.2.2 Discourse environments.

Next, we consider environmental distribution from the perspective of text genre. Texts comprising COHA are classified in terms of genre. Four such distinctions are made: fiction, magazines, newspapers, and non-fiction. While clearly coarse distinctions, they provide an important perspective from which to consider environmental distribution. Genre classifications are ultimately based in the intended audience of a particular text. In terms of the speech community, then, an audience can loosely be defined as a “discourse community” (Swales 1990). As Bizzel (1992:89) notes, “producing text within a discourse community cannot take place ... unless the writer can define her goals in terms of the community's interpretive conventions.” Ultimately, we want to capture the extent to which the usage of a given  $w$  is conventional across different discourse communities.

A vector-space model is presented here to approximate how lexica vary with respect to their distribution across genre,  $E$ . A word-environment matrix was constructed, in which the overall frequency of each  $w$  is represented in terms of its relative frequencies across each genre,  $E_i$ . Instead of comparing individual word vectors to assess distributional similarity (as in traditional VSM applications per §2), here we

compare each word vector to a single vector that represents maximal variation in  $\mathbf{E}$ . For our purposes here, maximal variation in genre distribution is assumed to be the relative frequency of the overall corpus,  $\mathbf{q}$ , across each genre,  $\mathbf{E}_i$ .

Table 3.3 summarizes the  $\mathbf{q} \mathbf{x} \mathbf{E}$  distribution for generation  $t_7$ . This distribution, then, reflects the relative prominence of each  $\mathbf{E}_i$  within the speech community. From this perspective, it could be said that 48% of our experiences in the speech community involve language categorized as fiction, for example. A given  $\mathbf{w}$  is maximally variable with respect to  $\mathbf{E}$  if its relative frequencies across  $\mathbf{E}_i$  equal the relative frequencies of  $\mathbf{E}_i$  in the speech community. In other words, to be maximally variable with respect to  $\mathbf{E}$  is to be used uniformly across  $\mathbf{E}_i$ .

*Table 3.3: Relative frequency distribution across genre for the corpus at  $t_7$ .*

Fiction	Magazines	News	Non-fiction
48%	27%	14%	11%

Table 3.4 presents  $\mathbf{w} \mathbf{x} \mathbf{E}$  vectors for two word forms from  $t_7$ : *chaos\_n* and *hypothesis\_n*. As can be noted, the  $\mathbf{w} \mathbf{x} \mathbf{E}$  vector for *chaos\_n* is essentially the same as  $\mathbf{q} \mathbf{x} \mathbf{E}$  in Table 3.4. The  $\mathbf{w} \mathbf{x} \mathbf{E}$  vector for *hypothesis\_n*, on the other hand, is quite different from  $\mathbf{q} \mathbf{x} \mathbf{E}$ . This difference is most notable when we compare the relative prominence of non-fiction text in the speech community (11%) to the relative frequency of *hypothesis\_n* in texts classified as non-fiction (71%).

Table 3.4: Relative frequency distributions of *chaos\_n* and *hypothesis\_n* across genre at  $t_7$

	Fiction	Magazines	News	Non-fiction
<i>chaos_n</i>	49%	29%	12%	10%
<i>hypothesis_n</i>	7%	19%	3%	71%

Distance from the  $q \ x \ E$  distribution is construed here in terms of variability with respect to genre: the further the distance, the lower the degree of variability.

Additionally, we make the assumption that lower variability with respect to discourse environments implies fewer discourse environments. From this perspective, it could be said that *hypothesis\_n* occurs in fewer, less variable discourse environments in comparison to *chaos\_n*.

We follow McDonald & Shillcock (2001) in quantifying this distance using relative entropy, or Kullback–Leibler divergence. Relative entropy measures the distance between two probability distributions, and is based in a comparison between an observed distribution (the posterior distribution) and an expected distribution (the prior distribution). Here, the observed distribution is  $w \ x \ E$  for some  $w$  and the expected distribution is maximal variation, ie,  $q \ x \ E$ . The deviation between  $w \ x \ E$  and  $q \ x \ E$  is calculated for each  $E_i$  as follows:

$$(1) \ RE_{E_i|w} = P(E_i | w) * \log_2 ( P(E_i | w) / P(E_i | q) )$$

In the case of *hypothesis\_n*, where  $E_i$  is “non-fiction,”  $P(E_i|w)$  is the relative frequency of *hypothesis\_n* in texts classified as non-fiction, or 71% (Table 3.4), and  $P(E_i|q)$  is the relative prominence of text classified as non-fiction in the speech community, or 11% (Table 3.3). Relative entropy for  $w$ , then, is the summation of individual  $E_i$  deviations presented in (1):

$$(2) \quad RE_w = \sum_{i=1}^n P(E_i | w) * \log_2 ( P(E_i | w) / P(E_i | q) )$$

Based on (2), the relative entropy for *chaos\_n* is 0.004, while the relative entropy for *hypothesis\_n* is 1.57. The number of (or degree of variability across) discourse environments is operationalized as the inverse of relative entropy; in this way, higher values reflect greater variation.

### §3.2.3 Local environments.

Lastly, we consider environmental distribution from the perspective of local co-occurrence. Local co-occurrence in a corpus is generally defined as some  $n \times n$  window of words surrounding a target word in running text. The size of  $n$  varies in the literature, and is task dependent. Peirsman et al. (2008) have shown that VSMs based in the distributional hypothesis are most optimal in capturing semantic similarity with window sizes ranging from  $n=4$  to  $n=7$ . While the task here is different, a window size

of 5 is selected. Example (3) below presents an excerpt of text extracted from COHA; the target *anxiety\_n* is underlined while the 5x5 window of local co-occurrence is boldfaced.

- (3) “He detected **frozen waffles and the usual anxiety on their breath. Valdek walked** backwards through the canyon of cages...” (COHA, Ploughshares:Spring 2003:. Vol. 29, Iss. 1; pg. 131:).

While the 5:5 window presented in (3) amounts to ten word forms, only forms tagged as open-class and non-proper are considered informative for our purposes here; as lemma-PoS pairs, these include *frozen\_a*, *waffle\_n*, *usual\_a*, *breath\_n*, and *walk\_v*. In terms of a VSM, these forms can be construed as five different local environments,  $E_i$ , in which *anxiety\_n* occurred. From this perspective, *anxiety\_n* occurred in a total of 2,760 unique local environments at  $t_7$ ; Table 3.5 presents the ten most frequent of these environments.

Table 3.5: Ten most frequent local environments for *anxiety\_n* at  $t_7$ .

social_a	state_n	depression_n	feel_v	level_n	high_a	fear_n	say_v	separation_n	physique
48	47	47	41	38	37	35	34	30	29

Traditional applications of VSMs in which  $E$  is defined as local co-occurrence generally define the scope of  $E_i$  as the  $i$ -th most frequent words in a corpus, in which  $i$

can range from 500-2,000 (and presumably higher). Again, the goal of such approaches is to be able to compare  $w_1$  and  $w_2$  in a uniform way in order to assess similarity. The goal here, however, is only to understand variation in local co-occurrence in relation to maximal variation in  $E$ . Importantly, we want this characterization to be based in how speakers most commonly experience a given  $w$ .

The argument presented here is that traditional approaches are not best suited to provide this particular characterization. First, an  $E$  scoped as the  $i$ -th most frequent forms in a corpus can often misrepresent the qualitative nature of common local environments for a given  $w$ . Consider, for example, an  $E$  at  $t_7$  in which  $E_i$  is defined as the 2,000 most frequent forms during  $t_7$ . If we consider the 25 most frequent local environments of *anxiety\_n* during  $t_7$ , we find that 10 (or 40%) are not included in  $E_i$  (eg, *depression\_n* in Table 3.6). Such environments very much define how speakers experience *anxiety\_n* locally.

Additionally problematic with traditional approaches is the issue of sparsity across  $E_i$ ;  $E$  scoped as the  $i$ -th most frequent forms in a corpus is mostly a description of local environments in which a given  $w$  is not used. If we consider all forms occurring at greater than 0.25 ppm across all  $t$ , the median count of unique local environments is 205. Percentages of these co-occurrences happening “only once” and “twice or less” were calculated for the same set of forms; the median value for the former is 86.9% and the median value for the latter is 96.2%. The median scenario, then, is that  $w$  occurs in 205 unique local environments, and in only 13% (or, 100% - 86.9%) of these environments more than once.

For these reasons, we define  $E_t$  as the fifty most frequent local environments per  $w$ . Instead of some uniformly scoped  $E_t$ , then, this operationalization shifts focus to the qualitative idiosyncrasies of local environments across  $w$ . From this perspective, maximal variation is defined per  $w$ . Table 3.6 compares the relative frequency of *anxiety\_n* at  $t_7$  across the ten most frequent local environments in which it occurs (as  $w \times E$ ) to the relative prominence of these local environments in the speech community (as  $q \times E$ ).

*Table 3.6:  $w \times E$  and  $q \times E$  for the ten most frequent local environments of *anxiety\_n* at  $t_7$*

social_a	state_n	depression_n	feel_v	level_n	high_a	fear_n	say_v	separation_n	physique_n
4.3%	4.2%	4.2%	3.7%	3.4%	3.3%	3.1%	3.0%	2.7%	2.6%
0.6%	1.4%	0.1%	2.7%	0.6%	1.7%	0.3%	16.4%	0.1%	0.1%

Relative entropy is derived from the fifty most frequent local environments per  $w$  following equations (1) and (2). Similar to the derivation of discourse environments, the number of (or degree of variability across) local environments is operationalized as the inverse of relative entropy; in this way, higher values reflect greater variation. Per McDonald & Schillcock (2001), local environment approximations are limited to forms occurring in at least 50 unique local environments per  $t$ ; the authors argue that estimates based on fewer unique local environments become less reliable.



### §3.2.4 Word length in syllables and part of speech.

While frequency, dissemination, discourse environments, and local environments are time-variant features of distribution, part of speech and word length in syllables do not vary across *t*. Table 3.7 summarizes the core lexicon by part of speech. As can be noted, the core lexicon is predominantly nominal: 55.3% (11,214/20,262) of the core lexicon is comprised of noun forms.

*Table 3.7: Core lexicon by part of speech.*

adjective	noun	verb	adverb
4,463	11,214	4,284	301

Word length in syllables is less straightforwardly a time-invariant variable; the relationship between frequency and phonetic reduction historically has been well-documented in grammaticalization research. However, from a corpus-based, dictionary-based perspective, detecting changes in word length over the last 200 years is virtually impossible. While reductions may be happening in actual speech, orthographic conventions are clearly more conservative with respect to change. So, while word length is time-variant, for our purposes here it is treated as time-invariant. Syllable counts, then, were extracted from the Webster's 1913 dictionary.

### §3.2.5 Summary: Features of distributional variation.

Values for frequency, dissemination, discourse environments, and local environments are observed in COHA across all  $t$ . Syllable counts and part of speech are observed across all  $t$  as well but are time-invariant. All variables are log-transformed to address issues of skewness.

The features of environmental distribution presented here are designed to capture distributional variation across three distinct linguistic environments within the speech community: text environments, discourse environments, and local environments. Ultimately, the three metrics reflect three perspectives from which to understand frequency distributions. Importantly, we make no assumptions about how variation in environmental distribution relates to the linguistic notion of “context of usage.” That said, the cline presented in (4) illustrates the relatedness of the four metrics of distribution considered in this study in terms of their theoretical proximity to word meaning.

(4) Frequency < Dissemination < Discourse environments < Local environments

From this perspective, local environments are theoretically the closest to word meaning while pure frequency counts are the furthest from word meaning.

The cline presented in (4) can also be interpreted in terms of scale within the speech community. \*Macro to micro.\* Frequency counts ultimately reflect the absolute prevalence of a given  $w$  in a speech community. Dissemination steps down in scale, and

as a metric is meant to capture prevalence of usage in terms of the number of speakers that use a given  $w$ . The discourse environment metric provides an even finer-grained perspective, and attempts to capture prevalence of usage across the discourse communities in which speakers use a given  $w$ . Lastly, local environments as a metric is meant to capture this prevalence as it is reflected in local patterns of co-occurrence.

As has been noted frequently in the literature, features of lexical variation are highly correlated, especially features derived from counts in a corpus. The approaches to environmental distribution presented here are (in theory) designed to capture variation in a corpus independent of frequency. A correlation matrix for features of distributional variation is presented in Table 3.8.

*Table 3.8: Correlations (Pearson's  $r$ ) among features of distributional variation historically.*

	(2)	(3)	(4)	(5)	(6)
(1) Dissemination	0.47***	0.39***	0.08***	0.039	0.07***
(2) Local environments	-	0.29***	0.19***	-0.02	0.08***
(3) Discourse environments	-	-	0.07**	-0.17***	0.03
(4) Frequency	-	-	-	-0.18***	0.35***
(5) Syllables	-	-	-	-	-0.12***
(6) Degree of polysemy	-	-	-	-	-

Despite a relatively high degree of inter-relatedness among this set of features of distributional variation, collinearity among the variables is fairly low. Table 3.10

summarizes condition values ( $\kappa$ ) by generation. The first row presents condition values when degree of polysemy is not included in the test; the second row presents condition values when degree of polysemy is included in the test as well. Baayen (2008) notes that medium (ie, non-harmful) collinearity is reflected in condition values around 15; all condition values in Table 3.9 are well below 15. This lack of problematic collinearity speaks to the efficacy of the approaches to environmental distribution presented here in capturing variation independent of frequency.

*Table 3.9: Collinearity ( $\kappa$ ) by generation.*

	$t_1$	$t_4$	$t_7$	Total
Without polysemy	5.86	6.45	6.89	6.14
With polysemy	6.69	7.50	8.48	7.24

### §3.3 A model of polysemy.

Having operationalized degree of polysemy and features of distribution historically, this section investigates the synchronic consequences of polysemy in a speech community from an historical perspective.

### §3.3.1 Methods.

Analysis is limited to generations for which we have observations of degree of polysemy (ie,  $t_1$ ,  $t_4$ , and  $t_7$ ) and to forms that are fully describable within each generation as summarized in Table 3.10. While the core lexicon is comprised of approximately 20 thousand forms, not all forms meet the criterion of occurring in at least 50 unique local environments during each generation. In sum,  $n=43,460$  forms (by generation) are fully describable.

*Table 3.10: Word forms by generation.*

Generation	Forms (n)
$t_1$	13,663
$t_4$	14,874
$t_7$	14,923
Total	43,460

Multiple linear regression is utilized to describe how a unit (or degree) of meaning is reflected in a speech community as approximated by features of distributional variation. From this perspective, the dependent variable is the degree of polysemy and the independent variables are frequency, dissemination, discourse environments, and local environments. The time-invariant feature of word length in syllables is additionally

included as an independent variable. Lastly, the model controls for both part-of-speech and generation, and their respective effects on the model are not included here.

Based on previous findings in the literature, we expect a unit (or degree) of meaning to be reflected in all features of distributional variation, ie, we expect higher degrees of polysemy to be reflected in higher degrees of frequency, dissemination, discourse environments, and local environments historically. We also expect higher degrees of polysemy to be associated with shorter word forms.

### **§3.3.2 Results.**

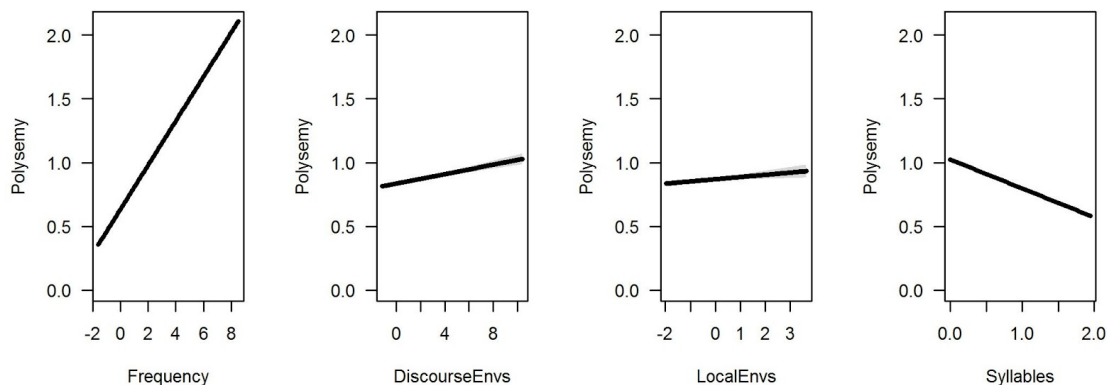
Results of the model are summarized in Table 3.11. Results align well with previous findings from the literature. As degree of polysemy increases in the mental lexicon, so to do frequency, discourse environments, and local environments in the speech community; degree of polysemy increases in the mental lexicon are not reflected, however, in increased degrees of dissemination within the speech community.

*Table 3.11: Summary of model of polysemy.*

	Coefficient	Std. Error	t value	Pr(> t )
(Intercept)	0.548904	0.011156	49.202	< 0.0001
FREQUENCY	0.173071	0.002006	86.266	< 0.0001
DISSEMINATION	-0.01028	0.008911	-1.153	0.24875
DISCOURSE ENVIRONMENTS	0.01855	0.002443	7.592	< 0.0001
LOCAL ENVIRONMENTS	0.01725	0.006296	2.74	0.00614
SYLLABLES	-0.22631	0.006001	-37.713	< 0.0001

A unit of meaning, then, is reliably reflected synchronically in units of frequency, discourse environments, and local environments; importantly, these relationships persist historically. Figure 3.1 presents an overview of the partial effects for significant predictors in the model of polysemy; as the figure illustrates (and t-values suggest), the effect size of frequency is by far the largest, followed by word length in syllables.

Figure 3.1: Partial effects of the significant predictors in the multiple regression model fitted to degree of polysemy. 95% intervals represented in gray.



While the model is based in historical data, the model itself is purely synchronic in nature. The fact that relationships between distributional variation and degree of polysemy persist historically is indeed interesting. This finding has not previously been reported in the literature; it ultimately demonstrates the temporal stability of these relationships, despite historical semasiological developments in the core lexicon (per §2). This type of stability is predicted by system-based accounts of the inter-relatedness of features of lexical variation (eg, Grzybek 2015).

Importantly, however, the consistency of these relationships does not help account for the development of novel semasiological structure from the perspective of either innovation or propagation, as the model is strictly synchronic and makes no distinction between speaker and speech community. The stability only speaks to the uniformity in which a unit of meaning is reflected in distribution historically; in other words, frequency, discourse environments, and local environments provide a fairly reasonable composite of a “context of usage” in a speech community. As to whether this composite



provides insights into the distributional features relevant to speakers in the process of innovation or features relevant to the development of novel form-meaning pairs in the speech community is the focus of the remainder of this dissertation.

#### **§4. LEXICAL VARIATION, COGNITIVE REPRESENTATION, AND SPEAKER MOTIVATIONS.**

The theoretical accumulative model presented in §1 is based exclusively in economic motivations, and makes the prediction that if some feature of distribution facilitates lexical access, it will also facilitate innovation and the development of novel form-meaning pairs. In part, this model is based on the relatedness between characteristics of highly polysemous forms and characteristics of forms that facilitate lexical access; it is also, in part, based in proposed “exponential” models of lexical development. A model based in competing intention types, however, makes the prediction that speakers will additionally recruit informative and familiar forms in processes of innovation per the motivations underlying processes of metaphor and metonymy (ie, expressiveness and not being misunderstood).

However, in order to evaluate which model better fits the historical data, we need a better sense of how cognitive notions of informativity, familiarity, and accessibility are sourced in features of distributional variation. In other words, we cannot test the prediction that forms used in innovation are informative, for example, if we do not know what informativity looks like in a corpus.

While linguists reference these types of cognitive constructs often, they are not especially well-defined terms. Per §1, researchers have speculated on relationships between various facets of distributional and behavioral variation and notions of informativity and familiarity. In this chapter, however, we consider research outside the domain of speaker intention and lexical innovation to provide a more definitive

characterization of these relationships, and to establish a set of predictions for a competitive model of innovation based in features of lexical variation.

The first section of this chapter considers previous research relevant to such a characterization. The second section provides a brief description of the data set. The third section presents predicted characteristics of innovative utility based in an accumulative model utilizing simple linear correlation. The fourth section presents predicted characteristics of innovative utility based on speaker intentions utilizing factor analysis.

#### **§4.1 Behavioral data and facets of cognitive representation.**

Bridging the “methodological gap” between cognitive constructs and distributional variation ultimately means understanding how features of distribution relate to speaker behaviors and speakers’ subjective knowledge of the lexicon (Schmid 2014). It also means being able to disentangle the highly interrelated space that is lexical variation in general (Baayen et al. 2006). Several studies have demonstrated evidence of underlying correlation structure among features of lexical variation, independent of (or in addition to) how such features relate to response times in lexical decision.

Clark & Paivio (2004), for example, use factor analysis to investigate the correlative relationships among over 30 features of behavioral and distributional features of variation, and the cognitive constructs that underlie them. The authors identify four

major constructs: three that loosely align with accessibility, familiarity, and informativity, and a fourth they describe as relating to word form. We present a novel iteration of this type of analysis utilizing the features of distribution considered in this study, as well as three features of behavioral variation — response times in lexical decision, age of acquisition ratings, and concreteness ratings.

The three types of behavioral data were selected for several reasons. First, they are the most frequently utilized behavioral data in the psycholinguistic literature. Second, all three data types have been normed and made publicly available. Third, an argument can be made that they loosely approximate the cognitive constructs that we are trying to understand. Clark & Paivio (2004) demonstrate a relationship between the notion of familiarity and age of acquisition ratings. Additionally, informativity and concreteness ratings have been related in psycholinguistic research (Katz 1989). Lastly, response times in lexical decision are generally interpreted in terms of accessibility.

While not directed explicitly at the questions posed here, previous research investigating the inter-relatedness of features of behavioral data and features of distributional data shed some light on how we might expect the cognitive constructs of accessibility, informativity, and familiarity to manifest themselves in a corpus.

Several behavioral studies, for example, have demonstrated a relationship between informativity (via concreteness) and environmental distribution. Schwanenflugel & Shoben (1983), for example, find that speakers more easily associate concrete *w* with specific environments of usage in a subjective rating task they dub “context availability.” Additionally, findings from word association tasks, which in part

capture characteristics of a given  $w$ 's local environment (Aitchison 2012), demonstrate that associative strengths between cue (as  $w$ ) and responses are greater for more concrete  $w$  (de Groot 1989). In other words, there is less variation in the response types elicited by more concrete  $w$ .

In theory, less variation in response type could reflect less variation in the lexical environments in which speakers experience more concrete  $w$ . This lack of environmental variation could also account for the ease with which speakers associate more concrete  $w$  with specific lexical environments. Based on these findings, then, we may expect informativity to be sourced in less variable (ie, more concentrated) text, discourse, and local environments. A similar argument is presented in Fenk-Oczlon & Fenk (2010). For more or less the same reasons, we would expect informativity to be reflected in lower degrees of polysemy (Nerlich & Clarke 1989).

Familiarity (via age of acquisition ratings) may also be sourced in environmental distribution, namely discourse environments. Baayen et al. (2006) demonstrate a relationship between age of acquisition ratings and the ratio of written-spoken frequencies: words occurring more frequently in spoken genres are rated as being acquired earlier. Based on this finding, we might expect familiar forms to be used across a variety of discourse environments.

Additionally, Clark & Paivio (2004) demonstrate a strong relationship between age of acquisition ratings and word length, and describe shorter word length as a feature of familiarity. The authors also describe familiarity in terms of concreteness ratings, suggesting that such ratings may not be exclusively a proxy for informativity. de Groot

(1989:836) makes a similar argument, positing that “across subjects there is more common knowledge with respect to concrete than abstract words.” Lastly, while frequency is generally considered the domain of accessibility via the well-attested relationship between frequency and response times, Fenk-Oczlon & Fenk (2010) describe familiarity in terms of frequency as well.

Clearly, notions of accessibility, informativity, and familiarity are not mutually exclusive; similarly, we do not assume here mutual exclusivity with respect to the types of information that features of distribution capture or reflect. This section is meant only to generate a rough set of predictions for the types of variation that may encompass the cognitive constructs underlying processes of innovation.

#### **§4.2 Methods.**

Factor analysis is utilized in this study to evaluate constructs underlying correlations among six features of distributional variation (including word length) and three features of behavioral variation. As noted in Clark & Paivio (2005), the utility of factor analysis in comparison to clustering techniques is that it allows individual features of lexical variation to be “multi-dimensional” across factors/constructs; clustering techniques, on the other hand, force mutual exclusivity. Assuming multi-dimensionality would seem to make cognitive sense, as notions of accessibility, familiarity, and informativity are clearly not mutually exclusive.

Response times for approximately 40 thousand forms are made available via the English Lexicon Project (Balota et al. 2007). Concreteness norms for approximately 37 thousand forms are made available by Brysbaert et al. (2013). Kuperman et al. (2012) make available age of acquisition ratings for approximately 30 thousand forms. As the three behavioral data sets are contemporary data sources, they align with lexical variation in the core lexicon at generation  $t_7$ .

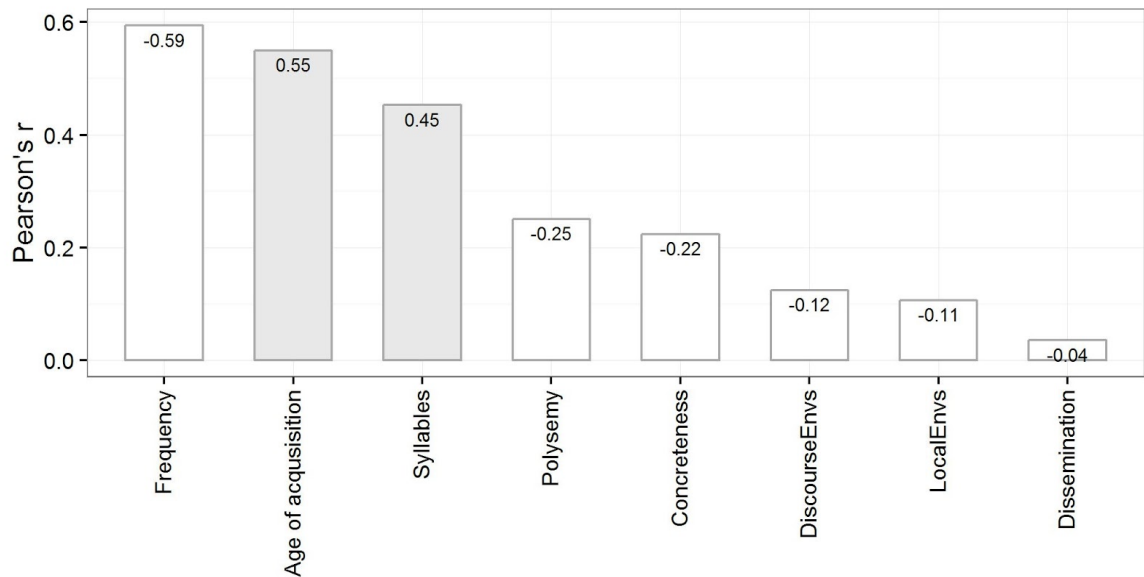
While the core lexicon can be described in terms of a lemma and part of speech, forms included in behavioral research tasks are generally only described in terms of the lemma. As a result, *gesture\_n* and *gesture\_v*, for example, are not distinguished in lexical decision and subjective rating tasks. For this reason, only forms that are non-ambiguous with respect to part of speech are included in the analysis. The set of forms matching this criterion in the core lexicon that are included in all three behavioral data sets amounts to 6,078 forms. Like the metrics of lexical variation, all three behavioral features are log-transformed to address skewness.

### **§4.3 Predictions based in an accumulative model.**

Figure 4.1 illustrates the magnitude (height) and direction (color) of the correlation between response times in lexical decision and each feature of lexical variation considered in this chapter, ordered by absolute magnitude of the correlation in terms of Pearson's  $r$ . While there is clearly variation in the magnitude of correlation, all

relationships are significant at  $p < 0.001$ . Importantly, the direction of each relationship is as predicted by previous research.

*Figure 4.1. Correlations between response times in lexical decision and features of lexical variation.*



Higher degrees of frequency and polysemy facilitate access, along with shorter word length. In terms of environmental distribution, broader usage across text, discourse, and local environments all facilitate lexical access as well. So, as discussed in §1, and confirmed here, distributional characteristics of polysemous word forms (§3) and the distributional characteristics of forms that facilitate lexical access are more or less the same.



#### **§ 4.4 Predictions based in competing motivations.**

Results from a factor analysis, however, demonstrate a more complex correlational structure than the unidimensional perspective presented in Figure 4.1. A total of nine features are considered (six distributional and three behavioral); all features were subjected to a principal components factor analysis with varimax rotation. A 3-factor solution was extracted, as three factors had eigenvalues greater than 1. The three factors account for approximately 50% of variation in the dataset.

Figure 4.2: Factorial structure among features of lexical variation.

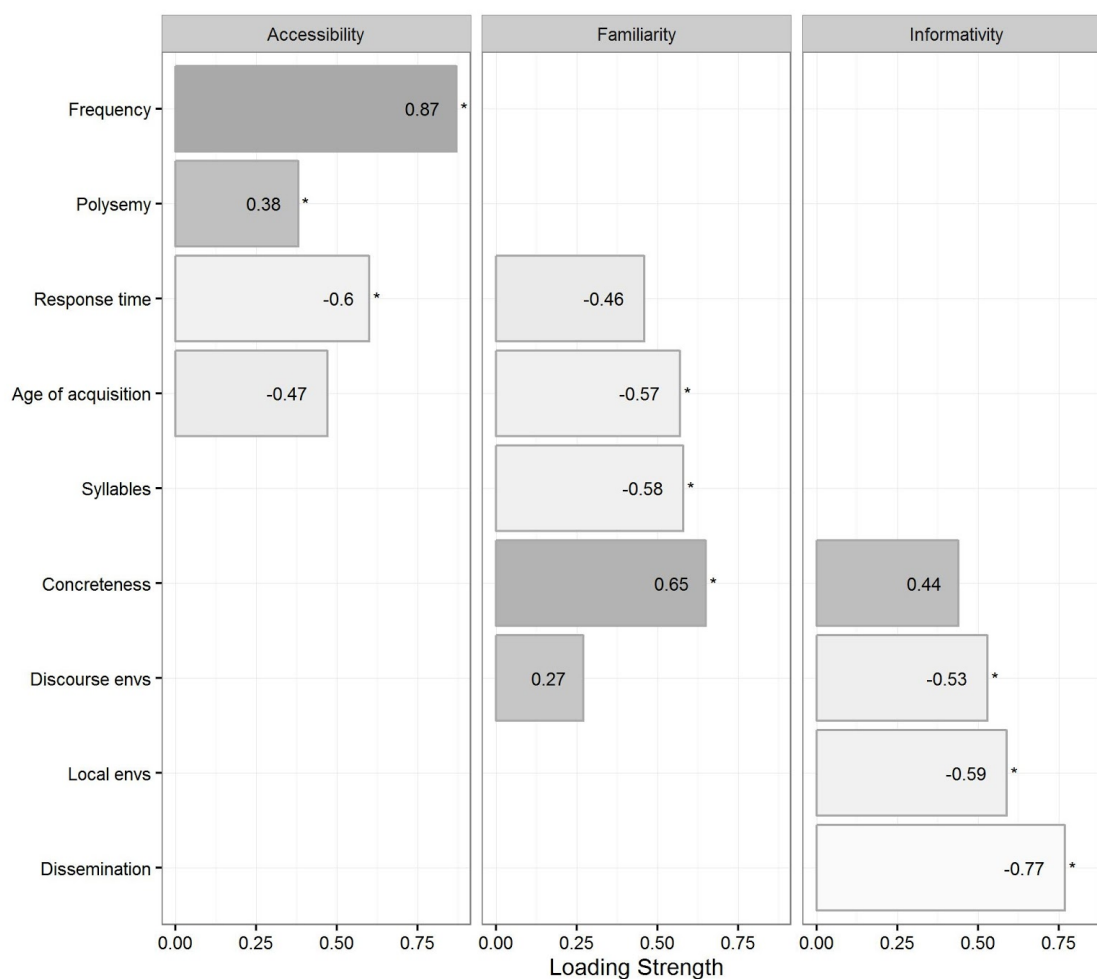


Figure 4.2 summarizes the factorial structure for each factor in terms of behavioral and distributional features. Only features with salient loadings (ie, greater than 0.25) are presented; the polarity of loadings can be interpreted as the direction of the correlation relative to other loadings within the same factor. As can be noted, some features load on more than one factor. The factor on which each feature loads the strongest is referred to here as the feature's primary loading. Primary loadings are denoted by an asterisk in Figure 4.2; secondary loadings are not. As Biber (1991:85)

notes, “a factor loading indicates ... the extent to which a given feature is representative of the dimension underlying a factor.”

The dimension (or cognitive construct) underlying Factor 1, then, is interpreted as accessibility. The primary loadings of frequency and response time straightforwardly support this interpretation. Additionally, degree of polysemy is a primary loading on this factor. While the expectation was that degree of polysemy relate to informativity, results demonstrate a strong association with accessibility.

Factor 3, on the other hand, is interpreted as informativity. All three metrics of environmental distribution load primarily on this factor, along with a secondary loading of concreteness. The relatedness of environmental distribution and concreteness ratings has been demonstrated previously in the literature. Informativity, then, is reflected behaviorally in higher concreteness ratings and distributionally in concentrated text environments, concentrated discourse environments, and concentrated local environments.

Lastly, the construct underlying Factor 2 is interpreted as familiarity. The primary loadings of concreteness, age of acquisition, and word length in syllables well align with the characterization of familiarity presented in Clark & Paivio (2004). The secondary loading of discourse environment additionally supports this classification, and suggests that there is a sociolinguistic component to the notion of familiarity. Familiarity, then, is reflected behaviorally in higher concreteness ratings and lower age of acquisition ratings, and distributionally in shorter word length and greater variation across discourse environments.

Following Clark & Paivio (2004), we assume conceptual similarities among the three factors are reflected in shared factor loadings. Response times and age of acquisition both load on Factors 1 and 2, for example. Additionally, concreteness loads on Factors 2 and 3; a similar pattern is attested in Paivio & Clark (2004)<sup>2</sup>. Both sets of similarities would seem to make intuitive sense. Perhaps most telling, however, is the absence of shared loadings between Factors 1 and 3; this finding suggests that notions of accessibility and informativity are distinct in nature. Lastly, it should be noted that shared factor loadings only occur with behavioral features of variation; distributional features load mutually exclusively across the three factors.

Linking distributional variation to behavioral variation via factor analysis not only provides insight into underlying constructs, it also provides a lens from which to understand the influence of cognitive representation historically on processes of lexical innovation historically, in the absence of historical behavioral data. As noted by Clark & Paivio (2004), experimentally collected subjective ratings change across generations of language users. If we assume that such ratings are sourced in linguistic experience, than this particular finding is ultimately predicted.

---

<sup>2</sup> While discourse environments load on Factors 2 and 3 as well, the polarities are not in the same direction.

## §5. SPEAKER MOTIVATIONS AND LEXICAL INNOVATION.

Having established relationships between distributional variation and speaker intentions via the cognitive constructs underlying speaker intentions in §4, we can now investigate the role of speaker intentions in processes of innovation historically. The semasiological development of the core lexicon attested in §2 demonstrates that speakers from different generations have been innovating historically. Ultimately, the goal of this chapter is to understand the role of speaker intentions in this development process. Functionally, we get at this distinction by comparing the features of forms at  $t$  that develop novel semasiological structure from  $t$  to  $t+1$  (ie, features of innovative utility), to the features of forms at  $t$  that do not develop novel semasiological structure from  $t$  to  $t+1$ .

Section 1 of this chapter presents a comparative dictionary operationalization of lexical development from  $t$  to  $t+1$ . It also describes the conceptual and methodological underpinnings of a temporally idealized model of lexical innovation in which the speaker and the speech community are treated distinctly; this model is compared to the model of polysemy presented in §3. Section 2 summarizes the predicted characteristics of innovative utility relative to speaker intention. Lastly, section 3 presents results from a logistic regression model.

## §5.1 Methods.

§2 presented some potential limitations of the semasiological distributional hypothesis, and the application of VSMs to the problem of identifying the development of novel form-meaning pairs historically, most notable being the issue of “false positives.” While a dictionary-based approach to the issue of novel semasiological structure has not previously been presented in the literature at scale, the findings presented in §2 provide strong support that the multiple dictionary resources utilized in this study are comparable from an aggregate distributional perspective. It was argued in §2 that such distributional uniformity supports an historical dictionary-based approach to the identification of novel form-meaning pairs at the word level.

Based on three historical observations of degree of polysemy (at  $t_1$ ,  $t_4$ , and  $t_7$ ), two historical comparisons can be made to assess whether a given  $w$  has developed novel meaning structure. Table 5.1 summarizes the degree of polysemy for six lexical units across generations  $t_1$ ,  $t_4$ , and  $t_7$ , as well as semasiological development between each historical comparison. Such developments are operationalized categorically as “no change,” “loss,” and “gain.”

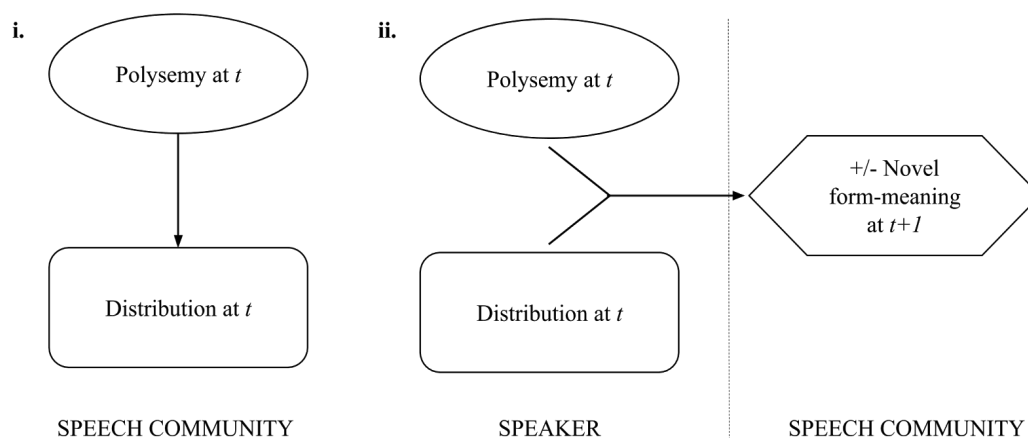
*Table 5.1: Polysemy and lexical development historically.*

Form	$t_1$	$t_4$	$\Delta(t_4 - t_1)$	$t_4$	$t_7$	$\Delta(t_7 - t_4)$
adulterate_v	2	1	loss	1	1	no change
anteroom_n	1	1	no change	1	1	no change
disappointment_n	1	2	gain	2	2	no change
handcuff_v	1	1	no change	1	2	gain
install_v	1	2	gain	2	3	gain
mundane_a	1	1	no change	1	2	gain

For modeling purposes, we generalize over these two “sets” of diachronic change as a single, two-step time series,  $t$  to  $t+1$ . As the comparison of interest here is between forms that develop novel form-meaning pairs from  $t$  to  $t+1$  and forms that do not develop novel form-meaning pairs from  $t$  to  $t+1$ , we consider only developments categorized as “gain” or “no change” in this analysis.

Based on this operationalization, novelty and degree of polysemy are treated distinctly; polysemy exists synchronically in  $t$ , while novelty develops (ie, propagates) diachronically between  $t$  and  $t+1$ . From this perspective, polysemy is just another feature of synchronic distribution. Importantly, the independent and diachronic treatment of novelty allows us to make a clear distinction between speaker and speech community — the speaker and her cognitive representation of each  $w$  are approximated via synchronic distribution, while change in the speech community is approximated diachronically via “gain/no change” in lexical development. Figure 5.1 illustrates the relationship between the model of innovation proposed here and the model of polysemy presented in §3.2.

Figure 5.1: A model of innovation in comparison to a model of polysemy.



This particular model design is conceptually akin to the experimental design presented in Katz (1989). Per §1, Katz (1989) presented speakers with an “innovation task,” and compared the characteristics of forms speakers recruited to complete the task to the characteristics of forms speakers did not recruit. Instead of a task, we take the development of novel semasiological structure from  $t$  to  $t+1$  as evidence that innovation occurred at  $t$ . Like Katz (1989), we assume that a characterization of forms at  $t$  that develop novel semasiological structure from  $t$  to  $t+1$  can provide insight into speaker intentions underlying processes of innovation historically.

The model presented here is clearly quite idealized. Per the architecture of the model, we assume that all semasiological development occurring in the mental lexicon historically is happening along the same timeline. In theory, novel form-meaning pairs that develop between  $t$  and  $t+1$  are sourced in innovations occurring at some



word-specific  $t'$ , which is defined here as the theoretical moment in time when an innovative usage of some  $w$  is introduced into the speech community. Ideally, a model of innovation would consider cognitive representation (via distribution in a corpus) at  $t'$  for each  $w$ . However, identifying  $t'$  systematically in a corpus is methodologically problematic; instead, we simplify the task by approximating  $t'$  as  $t$ .

In addition to the uniform treatment of  $t'$  as  $t$ , we also assume that cognitive representation is uniform across all speakers within a given speech community at some  $t$ . Clearly this is not the case — individual speakers experience concepts denoted in the lexicon in different ways; these different experiences result in “idio-representations.” Presumably these representations are the true basis for novelty in the lexicon historically. However, an aggregate perspective on cognitive representation is the best we can do utilizing COHA.

Lastly, it is important to note that while we have dubbed this model a model of “innovation,” it is more accurately a model of “successful innovation”; in other words, we only characterize forms used in innovation that become convention within a speech community as approximated by an additional dictionary entry. Forms used innovatively that do not become convention are not accounted for utilizing this particular methodology.

## §5.2 Predictions.

Table 5.2 summarizes the predicted characteristics of innovative utility by cognitive construct per the findings from the factor analysis in §4, as well as predictions based in theory from the literature. As noted in §4, predictions based in theory well align with the results of the factor analysis, with the exception of the role of polysemy. While several linguists have speculated on a relationship between polysemy and informativity, results from the factor analysis frame polysemy as principally a feature of accessibility. In the case of the former, then, the prediction is that less polysemous forms better serve the speaker intent of expressiveness; in the case of the latter, the prediction is that more polysemous forms better serve the speaker intent of being economical.

*Table 5.2: Predicted characteristics of forms used innovation per factor structure and theory.*

	Accessibility	Familiarity	Informativity
Polysemy	More polysemous	-	Less polysemous
Frequency	More frequent		-
Word length	-	Fewer syllables	-
Text environments	-	-	More concentrated
Discourse environments	-	Less concentrated	More concentrated
Local environments	-	-	More concentrated

Similarly, results from the factor analysis demonstrate the relatedness of discourse environments to both familiarity and informativity. As a feature of familiarity, the prediction is that less concentrated (ie, a broader range of) discourse environments better serve the speaker intent of not being misunderstood; on the other hand, as a feature of informativity, the prediction is that more concentrated discourse environments better serve the speaker intention of expressiveness.

In contrast, an accumulative model based in features of distribution that facilitate lexical access predicts that forms used in innovation will be more polysemous, more frequent, shorter in word length, and less concentrated across all three lexical environments. Despite the utility of the factor analysis in interpreting how features relate to speaker knowledge, the two models ultimately only make different predictions with respect to the features of environmental distribution.

In sum, we expect the speaker intent of economy to be evidenced by the recruitment of highly accessible forms in the lexicon. We expect the speaker intent of not being misunderstood to be evidenced by the recruitment of highly familiar forms in the lexicon. Lastly, we expect the speaker intent of expressiveness to be evidenced by the recruitment of highly informative forms in the lexicon

### §5.3 Model of lexical innovation.

The likelihood that a form develops novel semasiological structure from  $t$  to  $t+1$  given its distributional characteristics at  $t$  is modeled utilizing logistic regression.

Logistic regression is ultimately a classification technique; the goal here is to classify the lexicon historically in terms of whether or not forms develop novel meaning structure based on distributional characteristics at  $t$ . Independent variables, then, include degree of polysemy, frequency, word length, text environments, discourse environments, and local environments. Additionally included in the model are part of speech and generation. These controls are included to account for heterogeneity in the data sourced in part of speech and time; no predictions are made here with respect to how these variables influence the likelihood of innovation historically.

Analysis is limited to forms in the core lexicon that can be fully described from either  $t_1$  to  $t_4$  or  $t_4$  to  $t_7$ , and to forms that either develop novel semasiological structure from  $t$  to  $t+1$  or experience no change in semasiological structure from  $t$  to  $t+1$ . A total of  $n=18,821$  forms (by generation) meet these criteria. Table 5.3 summarizes these forms by change type and generation.

*Table 5.3: Distribution of change type by generation.*

$\Delta$ Type	$t_1$ to $t_4$	$t_4$ to $t_7$	$t$ to $t+1$
No Change	5,158	4,652	9,810
Gain	3,761	5,250	9,011
Total	8,919	9,902	18,821

Table 5.4 summarizes results of the model. Forms that develop novel form-meaning pairs from  $t$  to  $t+1$  are more frequent, shorter in word length, and less polysemous at  $t$  than forms that do not develop form-meaning pairs from  $t$  to  $t+1$ . Such forms also distribute in more concentrated discourse environments and more concentrated local environments (and to a lesser extent more concentrated text environments).

*Table 5.4: Summary of model of lexical innovation.*

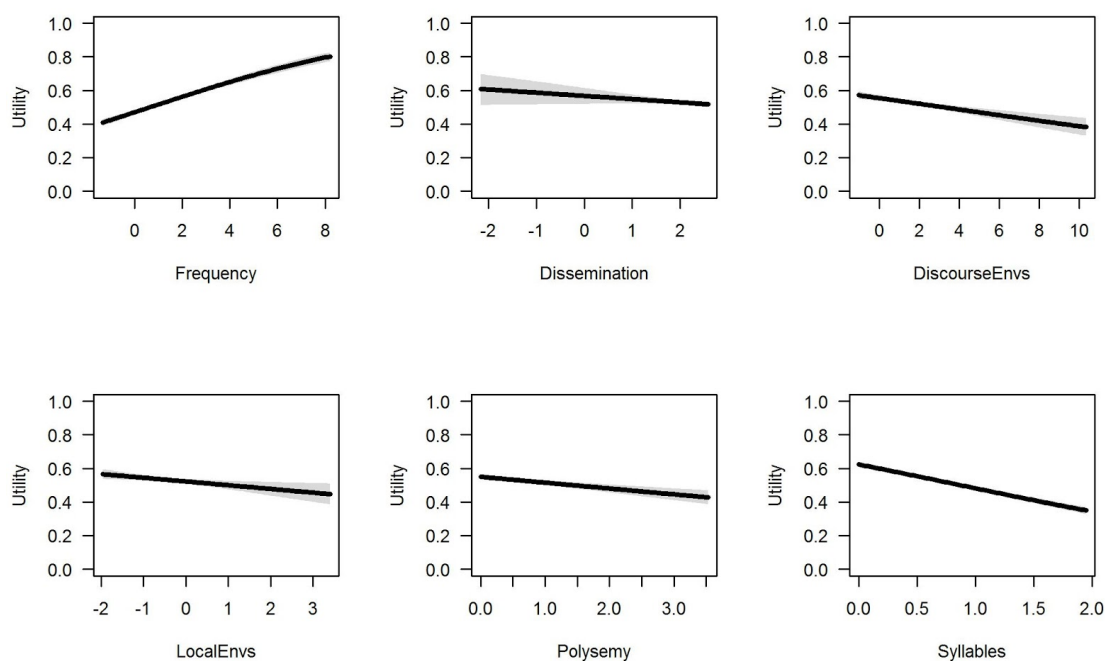
	Coef	S.E.	Wald Z	Pr(> Z )
(INTERCEPT)	-0.067	0.094	-0.71	0.476
FREQUENCY	0.180	0.014	12.8	<0.0001
DISSEMINATION	-0.086	0.047	-1.82	0.069
DISCOURSE ENVIRONMENTS	-0.070	0.013	-5.31	<0.0001
LOCAL ENVIRONMENTS	-0.097	0.035	-2.80	0.005
POLYSEMY	-0.139	0.028	-5.03	<0.0001
SYLLABLES	-0.592	0.033	-18.2	<0.0001

Results demonstrate that predictions based in speaker intention better align with model results than predictions based in an accumulative model. Divergence from the accumulative model is reflected in the direction of the relationship between innovative utility and the three metrics of environmental distribution. While an accumulative model predicts that forms distributing in more varied lexical environments are more likely to accrue novel semasiological structure historically, model results demonstrate the opposite.

Importantly, both sets of predictions make the wrong prediction regarding the direction of the effect of polysemy on innovative utility; despite the facilitative effects of degree of polysemy in lexical decision and proposed “exponential” models of lexical

development, innovative utility is sourced in lower degrees of polysemy historically. This effect is in line with the intuitions of Nerlich & Clarke (1989). The partial effects for each feature of distributional variation are presented in Figure 5.2. As can be noted, the effects of frequency and word length are the strongest; this finding mirrors results from the model of polysemy in §3.

*Figure 5.2: Partial effects of the significant predictors in the multiple regression model fitted to lexical development. 95% intervals represented in gray.*



Per the findings of §4, we interpret the direction of the relationship between innovative utility and the three metrics of environmental distribution as evidence for the role of expressiveness in processes of lexical innovation. As a feature of familiarity, we interpret the effect of word length on the model as evidence for the role of not being

misunderstood in processes of innovation. Lastly, as a feature of accessibility, we interpret the effect of frequency on the model as evidence for the role of economy in processes of innovation. Per partial effect sizes, an argument could be made for the prevalence of economy and not being misunderstood in processes of innovation historically.

As to how the effects of polysemy on the model should be interpreted in terms of speaker intention is less clear. While factor analysis demonstrates a strong relationship between accessibility and degree of polysemy, model results suggest that there are constraints on the influence of accessibility in processes of innovation with respect to polysemy that do not seem to exist for frequency. One explanation is that the speaker's need to be expressive trumps the need to be economical in the case of highly polysemous word forms; from this perspective, the effect of degree of polysemy on the model can be interpreted as additional evidence for the role of expressiveness in processes of innovation.

An alternative explanation is that the constraint is not speaker-based at all, but instead sourced in the hearer and speech community. Importantly, results from the model do not discount the possibility that speakers use highly polysemous forms in innovation; they only demonstrate that innovations based in less polysemous forms are more likely to become conventionalized usages in the speech community. From this perspective, it could be that hearers are less likely to repeat (ie, propagate) innovations based in highly polysemous forms because such innovations are not expressive. As Nerlich & Clarke

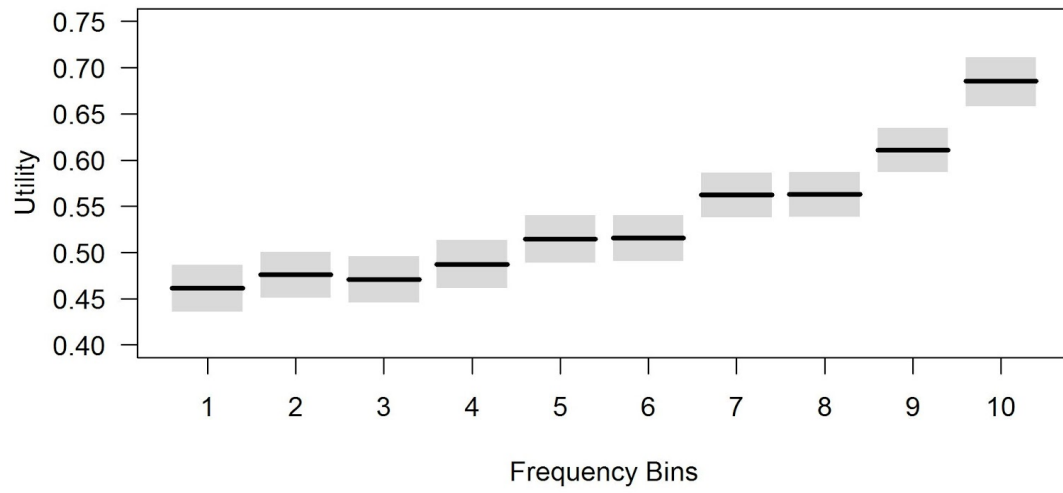


(1989:133) note, “how the hearer understands or (even more importantly) misunderstands a word or phrase will influence his/her future production.”

Along these lines, Clark & Gerrig (1983) investigate the role of degree of polysemy in the interpretability of lexical innovations, specifically eponymous verb phrases (eg, “Do a Napoleon for the camera”). While the experimental design is quite elaborate, suffice it to say, the authors include as a factor an experimentally constructed equivalent to monosemy, which they dub “coherency”; the authors found that the lack of coherency (ie, polysemy) detracts from the interpretability of these types of innovations. As to whether constraints on degree of polysemy in processes of lexical development are speaker- or hearer-based is ultimately an empirical question; however, regardless of who is responsible, it seems fair to suggest that at the root of the constraint is expressiveness.

Based on this line of reasoning, we consider potential constraints on accessibility in terms of frequency; while we assumed the relationship between frequency and innovative utility to be linear in the model presented in Table 5.4, it could be that at very high degrees of frequency the expressive utility of forms wanes. To investigate these potential nonlinear effects, we consider the effect of frequency across 10 (equally populated) frequency bins in a second iteration of the model; the partial effects of frequency as a categorical variable are summarized in Figure 5.2. As can be noted, effect-size increases more or less linearly from the least frequent forms in the core lexicon (bin 1) to the most frequent forms (bin 10); this finding suggests that there are not any constraints on accessibility in terms of frequency in processes of innovation as there are for degree of polysemy.

Figure 5.3: Partial effects by frequency bin.



## §6. INNOVATIVE UTILITY HISTORICALLY.

The focus of this chapter shifts away from processes of innovation occurring at  $t$  to the effects of the development of novel semasiological structure occurring between  $t$  and  $t+1$  in terms of distribution in the speech community. Per the idealized model presented in §5, such effects are assumed to reflect the aggregate effects of successful propagation in the speech community.

Ultimately, the interest here is how changes associated with this process influence the innovative utility of forms historically; more specifically, the interest is whether forms used in innovation at  $t$  are less likely to be used in innovation at  $t+1$ . While results from §5 shed important light on speaker intentions in processes of innovation, they do not entirely discount the viability of an accumulative model of lexical development historically. While the effect of degree of polysemy in the model of innovation suggests a higher degree of polysemy at  $t+1$  would necessarily detract from innovative utility at  $t+1$ , the strong linear relationship between frequency and innovative utility, as well as the “time-invariant” relationship between word length and innovative utility, suggest that an accumulative model may still hold diachronically when the net influences of the distributional features on innovative utility are considered in the aggregate.

The first section of this chapter, then, simply considers the effects of novel semasiological structure developed between  $t$  and  $t+1$  in terms of distributional change in the speech community. While several predictions from the literature are considered, this section is largely exploratory in nature. The second section, on the other hand, considers

how these changes affect innovative utility from  $t$  to  $t+1$ ; if, in fact, an accumulative model holds historically, we would expect forms that have developed novel semasiological structure from  $t$  to  $t+1$  to have higher degrees of innovative utility at  $t+1$  than at  $t$ .

### **§6.1 Consequences of propagation in the speech community.**

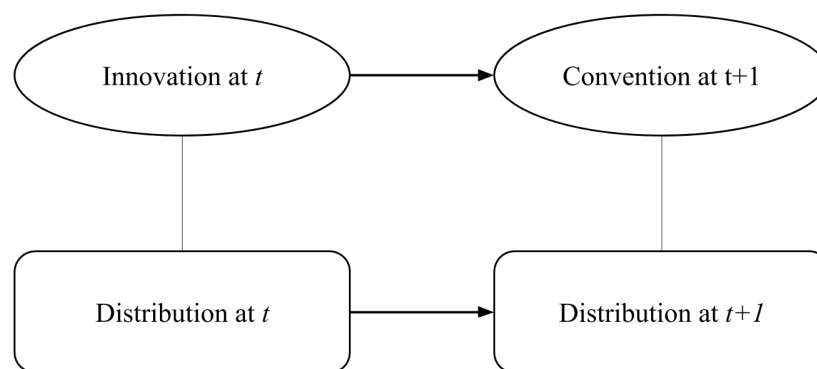
The model of polysemy presented in §3 considered the synchronic consequences of degree of polysemy in a speech community at  $t$ . This section, however, considers the diachronic consequences of the development of novel form-meaning pairs in a speech community from  $t$  to  $t+1$ . While much has been inferred regarding the effects of lexical development historically based on the synchronic relationships presented in §3, here these processes are explicitly investigated.

#### **§6.1.1 Methods.**

Panel analysis is utilized to capture the aggregate effects of processes of propagation in terms of changes in distribution from  $t$  to  $t+1$ . The goal of this particular regression analysis is to understand how changes in the dependent variable are reflected in changes in the independent variables across  $t$ . In comparison to other types of

longitudinal data (eg, time series data), panel data consist of many individuals and relatively few temporal points of observation (McManus 2011); as such, this method is well-suited for the data considered in this study. Figure 6.1 presents a schematic for the panel analysis model: degree of polysemy is treated as the dependent variable at  $t$  (the time of innovation) and  $t+1$  (the time of conventionalization); frequency, dissemination, discourse environments, and local environments are treated as the independent variables at  $t$  and  $t+1$ .

*Figure 6.1: Schematic for a diachronic relationship between processes of lexical development and changes in distributional variation.*



As we are only interested in the effects of the development of novel form-meaning pairs in terms of distribution in the speech community, we control for time-invariant lexical features (ie, word length and part of speech) using a fixed effects model; time-specific heterogeneity in the data is controlled for as well. Results from a

Hausman test support a time and entity fixed effects regression model (Torres-Reyna 2007).

It is predicted that propagative processes from  $t$  to  $t+1$  are reflected in the aggregate in higher degrees of dissemination in the speech community; while dissemination has (little to) no effect in the models of polysemy and innovation, as a text-based proxy for usership, increased degree of dissemination effectively is successful propagation. Altman et al. (2009) and Chesley & Baayen (2010) have demonstrated that the success and persistence of neologisms and lexical borrowings, respectively, is facilitated by the development of usership as reflected in dissemination values; functionally, a novel form-meaning pair is no different than a neologism or a borrowing.

From this perspective, it seems reasonable to expect the process of propagation to have an effect on discourse environments as well; presumably novel users take novel form-meaning pairs to novel discourse communities. Such a process additionally implies increased frequency in the speech community at-large. Lastly, we would expect the propagative process to affect local environments; ultimately, such changes (ie, novel patterns of local co-occurrence) are what lexicographers are responding to when adding entries to dictionaries (Kilgarriff 1997).

### §6.1.2 Results.

Analysis is limited to forms included in the model of innovation in §5 that develop novel semasiological structure from  $t$  to  $t+1$ . Table 6.1 summarizes panel analysis results; the effects of the development of a novel form-meaning pair in a speech community are increased frequency, increased dissemination, and a broader range of discourse environments. While dissemination is not relevant to either the model of polysemy or model of innovation, it relates strongly to the panel analysis, as predicted by findings presented in Altmann et al. (2009) and Chesley & Baayen (2010).

*Table 6.1: Summary of model of propagation.*

	Coefficient	Std. Error	t value	Pr(> t )
FREQUENCY	0.103163	0.003752	27.4986	<0.0001
DISSEMINATION	0.109419	0.008446	12.9548	<0.0001
DISCOURSE ENVIRONMENTS	0.014773	0.002619	5.6419	<0.0001
LOCAL ENVIRONMENTS	0.001588	0.007815	0.2032	0.839

The effects of novel semasiological structure, however, are not realized in local environments per model results. While we would expect novelty to influence patterns of local co-occurrence in the process of propagation, results would suggest that this is not the case historically. An alternative interpretation, however, is that changes in local

environments are not always indicative of the development of a novel form-meaning pair. As was discussed in §2, changes in local environments can additionally reflect changes in the overall composition of the lexicon or changes in content of the socio-cultural conversation; in other words, such changes are not always symptomatic of lexical development historically.

In the context of applying VSMS to the problem of lexical change, Gulordava & Baroni (2011:70) refer to such changes as “false-positives.” From this perspective, results of the model do not discount an influence of propagative processes on local environments; they only suggest that other factors influence variation in local environments independent of lexical development historically.

It should be noted here that the results of the panel analysis are importantly distinct from those presented in §3. Such results further demonstrate that synchronic relationships do not always capture the details of diachronic processes; inferring the impact of novelty within the speech community based exclusively on synchronic relationships would discount the role of dissemination in processes of propagation diachronically.

## **§6.2 Novel semasiological structure and innovative utility.**

An accumulative model of lexical development applied historically predicts that the development of a novel form-meaning pair from  $t$  to  $t+1$  for a given  $w$  makes  $w$  more



amenable to innovation at  $t+1$ . Results from the panel analysis provide evidence both for and against this type of model of lexical development. On one hand, the increase in frequency associated with lexical development theoretically increases the likelihood a form will be used innovatively by virtue of increased accessibility. On the other hand, increases in discourse environments and degree of polysemy theoretically detract from the informativity of a given  $w$ .

The question asked here, then, is what are the net effects of innovation on a given  $w$  in terms of innovative utility as described in the model of innovation presented in §5? If an accumulative model holds historically, we would expect forms that develop novel semasiological structure from  $t$  to  $t+1$  to have higher innovative utility at  $t+1$  than at  $t$ . We would also expect forms used innovatively at  $t$  to have higher degrees of innovative utility at  $t+1$  than forms at  $t+1$  not used in innovation at  $t$ .

### **§6.2.1 Methods.**

To investigate this particular question, the model of innovation presented in §5 is utilized to score innovative utility at the word-level historically. When applied in its predictive capacity, the output of a logistic regression model can be interpreted as a likelihood (or probability) that a given form will develop a novel form-meaning pair at  $t+1$  based on its distributional characteristics at  $t$ . While these likelihoods are based on known instantiations of change (and no change) from  $t_1$  to  $t_4$  and  $t_4$  to  $t_7$ , we can use the

historically trained model to predict the likelihood a given form will be used innovatively at any theoretical point in time. Table 6.2 presents innovative potential scores for four forms historically.

*Table 6.2: Innovative utility scores historically.*

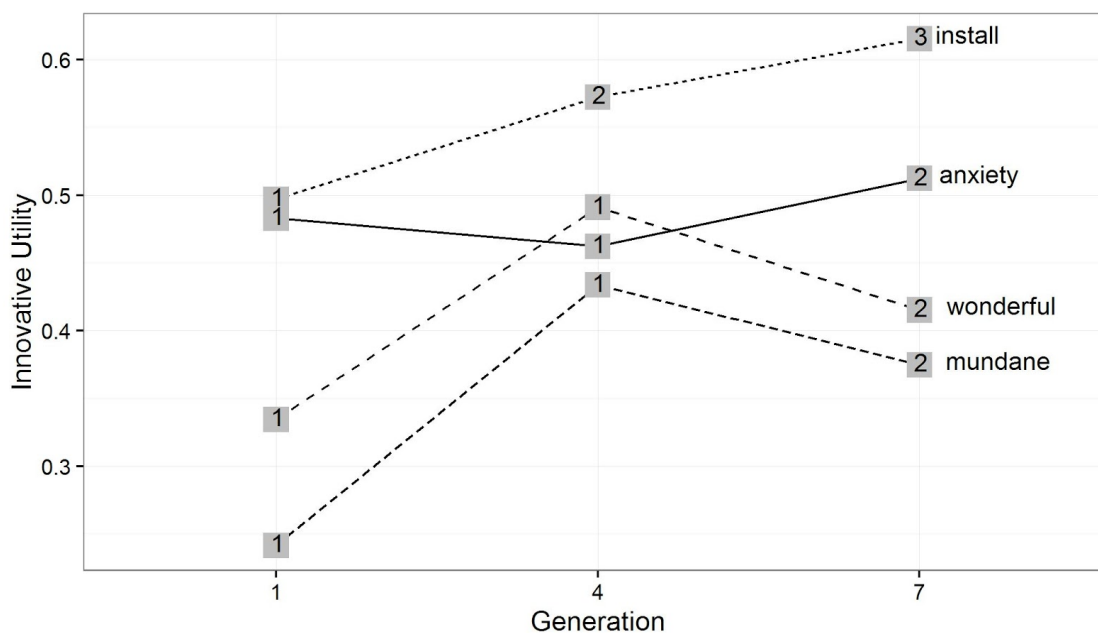
	$t_1$	$t_4$	$t_7$
<i>mundane_a</i>	0.24	0.43	0.37
<i>wonderful_a</i>	0.33	0.49	0.41
<i>anxiety_n</i>	0.48	0.46	0.51
<i>install_v</i>	0.49	0.57	0.61

Figure 6.2 illustrates the changes in innovative potential presented in Table 6.2 in relationship to the degree of polysemy for each of the four forms historically; values in gray boxes reflect the degree of polysemy while y-axis values reflect the likelihood of innovation. As can be noted from the figure, all four forms develop novel semasiological structure from  $t_4$  to  $t_7$ .

In the cases of *mundane\_a* and *wonderful\_a*, these developments result in losses in innovative potential; however, in the cases of *install\_v* and *anxiety\_n*, these developments result in increased innovative potential, despite the fact that additional degrees of polysemy necessarily detract from from such potential. In these cases,

presumably other changes in distribution (as consequences of the novelty) outweigh the effects of additional semasiological structure.

Figure 6.2: Innovative utility relative to semasiological structure by generation.



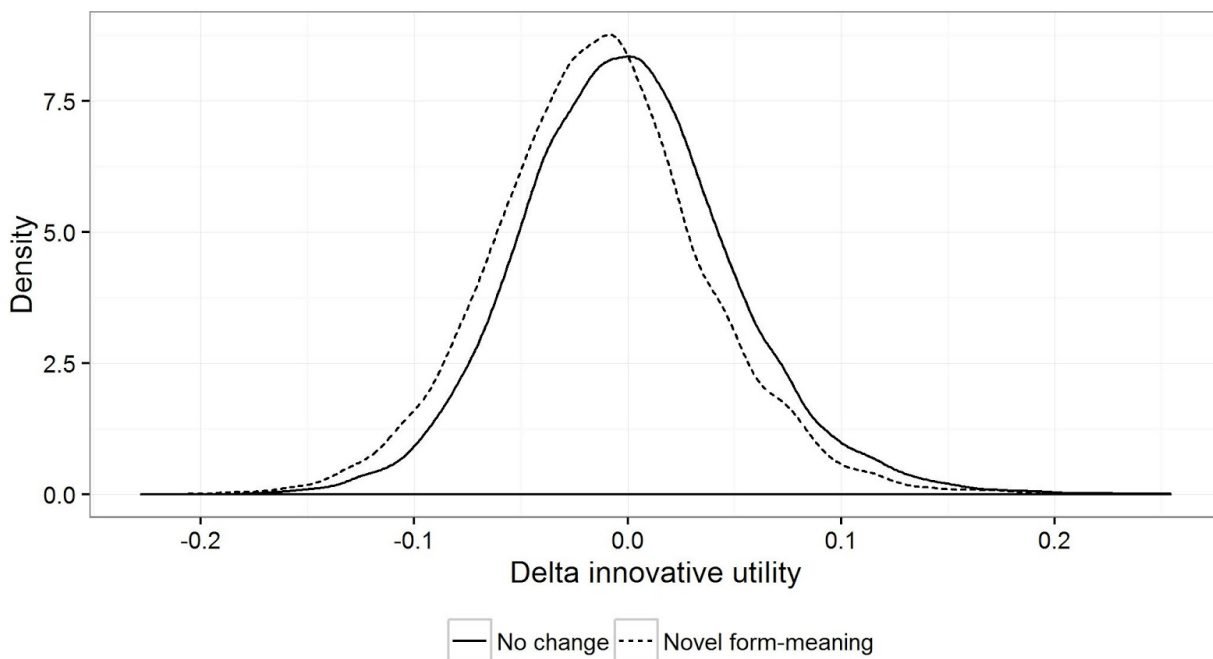
### §6.2.2 Results.

The effect of the development of novel semasiological structure on innovative potential across all lexica is investigated from two perspectives. First, we compare mean innovative potential scores at  $t$  to mean innovative potential scores at  $t+1$  for forms that have developed a novel form-meaning pair from  $t$  to  $t+1$ . Second, we compare changes in innovative potential between  $t$  and  $t+1$  as a function of whether or not forms have developed novel semasiological structure from  $t$  to  $t+1$ . Both analyses are conducted using paired Student's t-tests. Results from the first analysis demonstrate that innovative

potential scores are on average significantly lower at  $t+1$  than at  $t$  for forms that have developed a novel form-meaning pair from  $t$  to  $t+1$  ( $t = -16.56$ ,  $df = 15,611$ ,  $p\text{-value} < 0.0001$ ).

Results from the second analysis demonstrate that forms that develop a novel form-meaning pair from  $t$  to  $t+1$  have lower degrees of innovative utility at  $t+1$  than forms at  $t+1$  that have not developed such novelty ( $t = 7.22$ ,  $df = 14,421$ ,  $p\text{-value} < 0.0001$ ). Figure 6.3 presents probability distributions for change in innovative potential as a function of this distinction. As can be noted, while the change distribution for forms that do not experience lexical development from  $t$  to  $t+1$  is more or less centered at zero, this distribution is skewed to the left of center for forms experiencing lexical development.

Figure 6.3: Probability density plots for changes in innovative utility from  $t$  to  $t+1$  by change type.



Both sets of findings provide evidence that the pattern attested in the examples of *wonderful\_a* and *mundane\_a* in Figure 6.2 is the more prevalent one in the data; innovation at  $t$  generally means less likelihood of innovation at  $t+1$ . In other words, in the process of developing novel semasiological structure, forms lose innovative potential. Per §6.1, while the development of a novel form-meaning pair results in higher degrees of accessibility in the subsequent generation, the losses to informativity via additional semasiological structure and usage in a broader set of discourse environments are greater. The net effects, then, are in the direction of loss. These results provide fairly strong evidence against a model of lexical development based exclusively in economy (ie, the

accumulative model) historically, and demonstrate the mediating effects of competing motivations underlying processes of innovation.

## **§7. CONCLUSION.**

This dissertation set out to explore the competing motivations underlying processes of lexical innovation historically. A novel methodology was introduced to investigate speakers' intentions synchronically independently from degree of polysemy synchronically and the development of novelty in a speech community diachronically.

### **§7.1 Degree of polysemy historically.**

Findings from §2 demonstrated that generations of speakers over the last two centuries have been re-purposing members of the core lexicon to create novel form-meaning pairs. While monosemous forms comprise approximately 50% of the core lexicon in the 1828 publication of Webster's dictionary, such forms only constitute approximately 30% of the core lexicon in 2014. We demonstrated that this development is likely not a function of changes in lexicographic practices historically; probability distributions of degree of polysemy are essentially identical across random, independent samples of commonly used forms in the lexicon during each decade of publication.

Findings from §3 demonstrated that despite these developments historically, degree of polysemy is reflected in rather consistent ways in terms of features of distributional variation. While it has generally been speculated that the relationship between degree of polysemy and such features persists historically, the findings from §3

provide direct evidence of such maintenance. While frequency and word length are the strongest reflections of degree of polysemy synchronically, both local environments and discourse environments are relevant to such a characterization as well. The argument was made that collectively such features provide a composite characterization of how a unit of meaning is reflected in a speech community.

The influence of local environments (and discourse environments) on the model provide some support for the semasiological distributional hypothesis introduced in §2; however, frequency remains as a far better predictor. Results demonstrate that meaning in a corpus with respect to environmental distribution is indeed nebulous, and support a multidimensional approach to approximating degree of polysemy in a corpus.

## **§7.2 Speaker motivations and lexical development historically.**

§5 considered the role of speaker motivations underlying these historical developments, as evidenced by the distributional characteristics of the forms re-purposed by speakers to create these novel form-meaning pairs. The factor analysis presented in §4 provided a perspective from which to evaluate such intentions historically by linking contemporary features of distributional and behavioral variation to the cognitive constructs underlying speaker intention. Results demonstrated that accessibility is sourced in frequency and degree of polysemy, that familiarity is sourced in word length



and broader variation across discourse environments, and that informativity is sourced in concentrated text, discourse, and local environments.

Results from §5 demonstrated that speaker motivations are not limited to economy when innovating; instead, results provided evidence for the additional roles of expressiveness and not being misunderstood, as predicted by theory of metaphor (and metonymy) research. Findings also discounted proposed “exponential” models of lexical development: innovative utility is sourced in forms with lower degrees of polysemy, not higher degrees as an exponential model predicts. While this finding additionally runs counter to predictions based in factor analysis, it was hypothesized in §5 that competing motivations detract speakers from recruiting such forms in innovation, or that hearers are less likely to adopt innovations based in highly polysemous forms due to potential issues with the interpretability of such innovations.

Importantly, the characteristics of innovative utility presented in §5 are distinct from both the predicted features of innovative utility based in an accumulative (ie, economy exclusive) model of innovation and the features of highly polysemous word forms presented in §3.

As the findings from §2 attest, high degrees of polysemy are rare in the lexicon historically; forms having a degree of polysemy of 5 or greater only comprise approximately 8% of the lexicon at any point in time historically (Figure 2.1). Instead, the majority of the mental lexicon is comprised of monosemous forms, and approximately 85% of forms have degrees of polysemy less than 4. As findings from §5 demonstrate, this is ultimately where lexical development is sourced. It would seem that

a focus on anomalous degrees of polysemy in the lexicon has influenced previous usage-based, corpus-based approaches to lexical development; such focus has resulted in an over-emphasis on speaker economy in processes of innovation.

### **§7.3 The effects of novelty historically.**

In §6.1, it was demonstrated that the distributional consequences of the development (ie, propagation) of novelty in the speech community are fairly consistent historically: forms become more frequent, more disseminated, and used across a broader range of discourse environments. While dissemination did not figure in either a model of polysemy or a model of innovations, it was argued that a role of dissemination is predicted by the implications of successful propagative processes, ie, more speakers using some *w* in a novel way.

Findings from §6.1 demonstrated that local environments do not change in predictable ways in this process. While perhaps surprising, similar findings have been presented previously in the literature (eg, Gulordava & Baroni 2011); it was argued that changes in local environments can reflect changes in the content of the socio-cultural conversation historically, not just lexical developments.

§6.2 demonstrated how the net effects of this process make forms less amenable for innovation from one generation of speakers to the next. While the increase in frequency associated with the development of novelty makes forms theoretically more

accessible historically, the increase in discourse environments (and to a lesser extent increases in dissemination) detract from a forms expressive utility. It was argued that these results demonstrate why a theoretical accumulative model of lexical development does not hold historically.

#### **§7.4 Methodological implications.**

Much of the argumentation presented in this dissertation has been methodological in nature. The questions posed here have not been new ones per se; we have simply approached old linguistic problems from slightly different methodological perspectives.

##### **§7.4.1 Environmental distribution.**

Results from the three models presented in this dissertation demonstrate important differences in the distributional features relevant to degree of polysemy synchronically, speaker intention synchronically, and the development of novelty in a speech community diachronically. While the effects of frequency are present across all models, what distinguishes one model from the next are the magnitude and direction of effects associated with the features of environmental distribution described in §3. Table 7.1

summarizes model results in terms of these effects; only significant predictors are presented.

*Table 7.1: Features of distribution by model.*

	Polysemy	Innovation	Propagation
Frequency	+	+	+
Dissemination			+
Discourse environments	+	-	+
Local environments	+	-	

As the table illustrates, the features of environmental distribution behave differently across the three models and, as has been argued in this dissertation, in predictable ways. While results from the factor analysis in §4 demonstrate their synchronic relatedness, as well as a collective relatedness to subjective concreteness ratings, they translate differently across the three stages of lexical development considered in this dissertation.

Discourse environments and local environments relate more to degree of polysemy synchronically as well as speaker intention synchronically; on the other hand, dissemination and discourse environments relate more to processes of change in the speech community. Ultimately, these patterns demonstrate that environmental variation in a corpus should not be treated generically from a methodological perspective; they also

demonstrate the utility of a scaled approach in describing how lexica distribute across different linguistic environments in a corpus.

#### **§7.4.2 Synchrony, diachrony, and history.**

The methodological challenges of teasing apart process and effect from synchronic relationships between degree of polysemy and distribution were discussed in §1. This dissertation presented a novel approach to addressing these issues. By methodologically disentangling the speaker from the speech community over time, we were able to demonstrate important differences in the distributional features relevant to degree of polysemy synchronically, speaker intention synchronically, and the development of novelty in a speech community diachronically that a strictly synchronic approach glosses over.

Findings from §4 demonstrated the utility of factor analysis in interpreting the cognitive constructs underlying the complex set of correlative relationships that exist among features of lexical variation. Including behavioral features in this analysis enabled a more confident interpretation of these constructs in terms of speakers' subjective knowledge of the lexicon. Importantly, the findings from §4 provided a cognitive perspective from which to view features of distributional variation historically.

### §7.4.3 Limitations.

As noted in §5, the models presented in this dissertation are temporally idealized. In reality, innovation is happening all the time, and innovation becomes convention at different rates historically. As such, the approach to lexical development presented in this dissertation glosses over word-specific histories of semasiological development that are presumably fantastically idiosyncratic.

Along these lines, the methodology presented here idealizes over speakers' representations and the speech community at-large. Individual speakers experience words differently and have different mental lexicons; similarly, the "community" of speakers of American English is far from homogeneous. In this sense, the models presented here are at best global approximations of processes driven by unique individuals in unique linguistic communities.

As to whether the findings described in this dissertation have cross-linguistic utility is unclear. Presumably speaker motivations in innovative processes are the same cross-linguistically; from a methodological perspective, however, how cognitive notions like accessibility, familiarity, and informativity are manifested distributionally may likely be language-specific. This is clearly an empirical question.

### **§7.5 Discussion.**

This dissertation investigated competing motivations underlying processes of lexical innovation historically. Two sets of motivations were considered, one based in economy and the synchronic relationships between degree of polysemy and usage, and one based in expressiveness/not being misunderstood and theory of metaphor research. The unified approach to the question of speaker intention demonstrated evidence for all three intention types in processes of innovation historically.

The temporally idealized model presented here enabled us to ask questions of historical data with respect to speaker intentions that we have not previously been able to. By disentangling speaker from speech community methodologically, we were able to demonstrate important differences in the distributional features relevant to degree of polysemy synchronically, speaker intention synchronically, and the development of novelty in a speech community diachronically.

## References.

- Adelman, J., Brown, G., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814-823.
- Aitchison, J. (2010). Disasters, devastation and polysemy. *The Mental Lexicon, 5*(2), 163-179.
- Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon*. John Wiley & Sons.
- Aitchison, J., & Lewis, D. M. (2003). Polysemy and bleaching. In B. Nerlich, Z. Todd, V. Herman, & D. Clarke (Eds.), *Polysemy: Flexible Patterns in the Mind* (pp. 253–265). Berlin: Mouton De Gruyter.
- Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2011). Niche as a determinant of word fate in online groups. *PloS one, 6*(5), e19009.
- Altmann, E. G., Whichard, Z. L., & Motter, A. E. (2013). Identifying trends in word frequency dynamics. *Journal of Statistical Physics, 1-12*.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
- Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language, 55*(2), 290-313.
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human perception and performance, 10*(3), 340.
- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior research methods, 39*(3), 445-459.
- Biber, D. (1991). *Variation across speech and writing*. Cambridge University Press.
- Bizzell, P. (1992). *Academic discourse and critical consciousness*. University of Pittsburgh Press.
- Blank, A. (1999). Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change. *Historical semantics and cognition, 13*(1), 6.



- Blank, A. (2003). Polysemy in the lexicon and in discourse. *Trends in linguistics studies and monographs*, 142, 267-296.
- Blumenthal-Dramé, A. (2012). *Entrenchment in usage-based theories: what corpus data do and do not reveal about the mind* (Vol. 83). Walter De Gruyter.
- Britton, B. K. (1978). Lexical ambiguity of words used in English text. *Behavior research methods & Instrumentation*, 10(1), 1-7.
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2013). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods*, 1-8.
- Bybee, J. (1985). *Morphology: A study of the relation between meaning and form* (Vol. 9). John Benjamins Publishing.
- Bybee, J. (2015). *Language change*. Cambridge University Press.
- Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press.
- Chesley, P., & Baayen, R.H. (2010) Predicting new words from newer words: Lexical borrowings in French. *Linguistics* 48, 1343-1374.
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1), 22-29.
- Clark, H. H., & Gerrig, R. J. (1983). Understanding old words with new meanings. *Journal of verbal learning and verbal behavior*, 22(5), 591-608.
- Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, 767-811.
- Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers*, 36(3), 371-383.
- Croft, W. (1998). Linguistic evidence and mental representations. *Cognitive linguistics*, 9, 151-174.
- Croft, W. (2000). *Explaining language change: An evolutionary approach*. Pearson Education.
- Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.
- Davies, Mark. (2010-) *The Corpus of Historical American English: 400 million words, 1810-2009*. Available online at <http://corpus.byu.edu/coha/>.

- De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, 40(1), 213-231.
- De Dyne, S., & Storms, G. (2015). Word associations. In John R. Taylor (Ed.), *The Oxford Handbook of the Word* (pp.465-480). Oxford: Oxford University Press.
- de Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 824.
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological methods*, 5(2), 155.
- Ellis, N.C. (2013). Emergentism. In C. Chapelle (Ed.), *Encyclopedia of Applied Linguistics*, vol. 3 (pp.1873-82). Oxford: Wiley-Blackwell.
- Erk, K. (2012). Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass*, 6(10), 635-653.
- Fellbaum, C. (1998). *WordNet*. Blackwell Publishing Ltd.
- Fenk-Oczlon, G. (2001). Familiarity, information flow, and linguistic form. *Typological Studies in Language*, 45, 431-448.
- Fenk-Oczlon, G. & Fenk, A. (2010). Frequency Effects on the Emergence of Polysemy and Homophony. *Information Technologies & Knowledge*, 103-109.
- G. van Hell, J., & De Groot, A. M. (1998). Disentangling Context Availability and Concreteness in Lexical Decision and Word Translation. *The Quarterly Journal of Experimental Psychology: Section A*, 51(1), 41-63.
- Geeraerts, D. (1983). Reclassifying semantic change. *Quaderni di semantica*, 4, 217-240.
- Geeraerts, D. (1997). Diachronic prototype semantics. *A contribution to historical lexicology*.
- Geeraerts, D. (2015). How words and vocabularies change. In John R. Taylor (Ed.), *The Oxford Handbook of the Word* (pp.416-430). Oxford: Oxford University Press.
- Gilquin, G., & Gries, S. T. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory*, 5(1), 1-26.
- Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International journal of corpus linguistics*, 13(4), 403-437.

- Grzybek, Peter. (2015). Word length. In John R. Taylor (Ed.), *The Oxford Handbook of the Word* (pp.89-119). Oxford: Oxford University Press.
- Gulordava, K., & Baroni, M. (2011, July). A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. In *Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics* (pp. 67-71). Association for Computational Linguistics.
- Hanks, P. (2000). Do word meanings exist?. *Computers and the Humanities*, 34(1), 205-215.
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3), 146-162.
- Haspelmath, M., Calude, A., Spagnol, M., Narrog, H., & Bamyaci, E. (2014). Coding causal–noncausal verb alternations: A form–frequency correspondence explanation. *Journal of Linguistics*, 50(03), 587-625.
- Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics*, 39 (6; ISSU 376), 1041-1070.
- Hay, J. (2003). *Causes and consequences of word structure*. Psychology Press.
- Hay, J., & Baayen, H. (2002). Parsing and productivity. In *Yearbook of Morphology 2001* (pp. 203-235). Springer Netherlands.
- Heine, B., Claudi, U., & Hünnemeyer, F. (1991). *Grammaticalization: A conceptual framework*. University of Chicago Press.
- Hills, T. T., & Adelman, J. S. (2015). Recent evolution of learnability in American English from 1800 to 2000. *Cognition*, 143, 87-92.
- Hilpert, M., & Gries, S. T. (2009). Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing*, 24(4), 385-401.
- Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior research methods*, 45(3), 718-730.
- Jamrozik, A., Sagi, E., Goldwater, M., & Gentner, D. (2013, June). Relational words have high metaphoric potential. In *Proceedings of the 2013 Meeting of the North American Association for Computational Linguistics: Human Language Technologies, First Workshop on Metaphor in NLP* (pp. 21-26).

- Jurafsky, D., & Martin, J. H. (2000). *Speech & language processing*. Pearson Education India.
- Katz, A. N. (1989). On choosing the vehicles of metaphors: Referential concreteness, semantic distances, and individual differences. *Journal of Memory and language*, 28(4), 486-499.
- Kilgarriff, A. (1993). Dictionary word sense distinctions: An enquiry into their nature. *Computers and the Humanities*, 26(5-6), 365-387.
- Kilgarriff, A. (1997). I don't believe in word senses. *Computers and the Humanities*, 31(2), 91-113.
- Kintsch, W. (2007). Meaning in context. In T. K.Landauer, D.McNamara, S.Dennis, & W.Kintsch (Eds.), *Latent semantic analysis* (pp.89–105). Mahwah, NJ: Erlbaum.
- Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language*, 45(2), 259-282.
- Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, 20(1), 1-24.
- Köhler, R. (1986). *Zur linguistischen Synergetik: Struktur und Dynamik der Lexik* (Vol. 31). Bochum: Brockmeyer.
- Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *Journal of Experimental Psychology: General*, 140(1), 14.
- Kulkarni, V., Al-Rfou, R., Perozzi, B., & Skiena, S. (2015, May). Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 625-635). International World Wide Web Conferences Steering Committee.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4), 978-990.
- Labov, W. (1994). *Principles of linguistic change, Vol. 1: Internal factors*. Cambridge, Blackwell, 1994.
- Ladd, D. R., Roberts, S. G., & Dediu, D. (2015). Correlational studies in typological and historical linguistics. *Annu. Rev. Linguist.*, 1(1), 221-241.
- Lakoff, G. & Johnson, M.(1980). *Metaphors we live by*. University of Chicago.

- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, *104*(2), 211.
- Lee, C. J. (1990). Some hypotheses concerning the evolution of polysemous words. *Journal of Psycholinguistic Research*, *19*(4), 211-219.
- McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*(3), 295-322.
- McManus, P. A. (2011, October). Introduction to regression models for panel data analysis. In *Workshop in methods, USA: Indiana University. Slides*.
- Merriam-Webster*. (2014). Merriam-Webster, n.d. Web. 1 Dec. 2014.
- Murphy, G. L. (1997). Polysemy and the creation of new word meanings. In T. B. Ward, S. M. Smith & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 235-265). Washington, DC: American Psychological Association.
- Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure?. *Memory & cognition*, *28*(6), 887-899.
- Nerlich, B., & Clarke, D. D. (1989). Outline of a model for semantic change. *Diachrony within Synchrony: Language History and Cognition*, 125-144.
- Nerlich, B., & Clarke, D.D. (1992). Semantic change: Case studies based on traditional and cognitive semantics. *Journal of Literary Semantics* *20*(3): 204--225.
- Nerlich, B., & Clarke, D. D. (2001). Ambiguities we live by: Towards a pragmatics of polysemy. *Journal of Pragmatics*, *33*(1), 1-20.
- New, B. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic bulletin & review*, *13*(1), 45-52.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.
- Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Mahwah, NJ: Erlbaum.

- Paivio, A., Clark, J. M., Digdon, N., & Bons, T. (1989). Referential processing: Reciprocity and correlates of naming and imaging. *Memory & Cognition*, 17(2), 163-174.
- Pavio, A., Yuille, J.C. and Madigan, S.A. (1968). *Concreteness, imagery and meaningfulness values for 925 words*. Journal of Experimental Psychology Monograph Supplement, 76 (3, part 2).
- Peirsman, Y., Heylen, K., & Geeraerts, D. (2008). Size matters: tight and loose context definitions in English word space models. In *Proceedings of the ESSLLI Workshop on Distributional Lexical Semantics* (pp. 34-41).
- Polikarpov, A. A. (2006). *Cognitive model of lexical system evolution and its verification*. Retrieved 15 May, 2015, from [http://www.philol.msu.ru/~lex/articles/cogn\\_ev.htm](http://www.philol.msu.ru/~lex/articles/cogn_ev.htm)
- Porter, N. (1913). *Webster's Revised Unabridged Dictionary of the English Language*. G. & C. Merriam Company. *Project Gutenberg. Web. 1 December 2014.*
- Pustet, R. (2004). Zipf and his heirs. *Language Sciences*, 26(1), 1-25.
- Pylkkänen, L., Llinás, R., & Murphy, G. L. (2006). The representation of polysemy: MEG evidence. *Journal of cognitive neuroscience*, 18(1), 97-109.
- Rapp, R. (2003). Word sense discovery based on sense descriptor dissimilarity. In *Proceedings of the ninth machine translation summit* (pp. 315-322).
- Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46(2), 245-266.
- Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, 28(1), 89-104.
- Sagi, E., Kaufmann, S., & Clark, B. (2011). Tracing semantic change with latent semantic analysis. *Current methods in historical semantics*, 73, 161.
- Schmid, H. J. (2010). Does frequency in text instantiate entrenchment in the cognitive system. *Quantitative methods in cognitive semantics: Corpus-driven approaches*, 101-133.

- Schmid, Hans-Jörg. (forthcoming). Introduction: a framework for understanding linguistic entrenchment and its psychological foundations in memory and automatization. *Entrenchment, Memory and Automaticity: The Psychology of Linguistic Knowledge and Language Learning*, ed. by Hans-Jörg Schmid, 1–24. Boston: APA/Walter de Gruyter.
- Schmid, H. J., & Mantlik, A. (2015). Entrenchment in historical corpora? Reconstructing dead authors' minds from their usage profiles. *Anglia*, 133(4), 583-623.
- Scholz, F. W., & Stephens, M. A. (1987). K-sample Anderson–Darling tests. *Journal of the American Statistical Association*, 82(399), 918-924.
- Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(1), 82.
- Sorell, Joseph. (2015). Word frequencies. In John R. Taylor (Ed.), *The Oxford Handbook of the Word* (pp.68-88). Oxford: Oxford University Press.
- Steyvers, M., & Tenenbaum, J. B. (2005). The Large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, 29(1), 41-78.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Tabak, W., Schreuder, R., & Baayen, R. H. (2005). Lexical statistics and lexical processing: semantic density, information complexity, sex, and irregularity in Dutch. *Linguistic evidence—Empirical, theoretical, and computational perspectives*, 529-555.
- Taylor, John R. (Ed.). (2015). *The Oxford Handbook of the Word*. Oxford: Oxford University Press.
- Torres-Reyna, O. (2007). Panel data analysis fixed and random effects using Stata (v. 4.2). *Data & Statistical Services, Princeton University*.
- Traugott, E. C. (1980). Meaning-change in the development of grammatical markers. *Language sciences*, 2(1), 44-61.
- Traugott, E. C., & Dasher, R. B. (2001). *Regularity in semantic change* (Vol. 97). Cambridge University Press.

- Traugott, E. C., & Trousdale, G. (2013). *Constructionalization and constructional changes* (Vol. 6). Oxford University Press.
- Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37(1), 141-188.
- Webster, N. (1828). *Noah Webster's first edition of an American dictionary of the English language*. Foundation for American Christian Education. *Project Gutenberg*. Web. 1 December 2014.
- Zipf, G. K. (1949). Human behavior and the principle of least effort.