

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Word learning in context : the role of lifetime language input and sentential context

Permalink

<https://escholarship.org/uc/item/0492z47z>

Author

Borovsky, Arielle

Publication Date

2008

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Word learning in context:

The role of lifetime language input and sentential context

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Cognitive Science

by

Arielle Borovsky

Committee in charge:

University of California, San Diego

Professor Jeff Elman, co-chair
Professor Marta Kutas, co-chair
Professor Gedeon Deak
Professor Rachel Mayberry
Professor Nina Dronkers

2008

©
Arielle Borovsky, 2008
All rights reserved

The dissertation of Arielle Borovsky is approved and is in acceptable in quality and form for publication on microfilm

Co-chair

Chair

University of California, San Diego

2008

TABLE OF CONTENTS

Signature Page.....	iii
Table of Contents	iv
List of Figures.....	ix
List of Tables.....	xiii
Acknowledgments	xiv
Vita	xvi
Abstract of the Dissertation	xvii
Chapter 1	1
Background and Significance.....	1
1.1 The effects of lifetime language experience on early lexical and cognitive development	4
1.1.1 Vocabulary spurt theories	7
1.1.1.1 Naming insight.....	8
1.1.1.2 Social/pragmatic development	9
1.1.1.3 Constraints & Principles	9
1.1.1.4 Connectionism.....	11
1.1.1.5 Categorization	14
1.1.2 Issues in cognitive and linguistic change.....	15
1.1.2.1 Continuous vs. Discontinuous	16
1.1.2.2 Domain-general vs. Domain Specific.....	18
1.1.3 Why simulations?.....	21
1.1.4 Measurement of category development	22
1.1.5 Main questions: Global experience	28
1.1.5.1 Effect of Quantity	28
1.1.5.2 Effect of Frequency	28
1.1.5.3 Effect of Syntax.....	29
1.2 Effects of local experience on word learning:	30
1.2.1 Event-related potentials	33
1.2.2 Neural correlates of word learning	37

1.2.3 Neural substrates of word learning	42
1.2.4 Hemispheric asymmetries in lexical representation.....	45
1.2.5 Contextual Support and Semantic Knowledge	48
Chapter 2.....	53
Language input and semantic categories: A relation between cognition and early word learning.....	53
2.1 Abstract	53
2.2 Introduction.....	54
2.2.1 Relationships between lexical Development and category knowledge	55
2.2.2 The Role of Input.....	57
2.3 The Modeling Task	61
2.3.1.1 Simple recurrent models.....	62
2.3.1.2 Measurement of category structure	62
2.3.2 Effect of Quantity	65
2.3.3 Effect of Frequency	67
2.3.4 Effect of Syntax	68
2.4 Experiment 1: Effect of Quantity.....	69
2.4.1 Method.....	69
2.4.1.1 Input	69
2.4.1.2 Training.....	70
2.4.1.3 Analysis.....	72
2.4.2 Results and Discussion	73
2.5 Experiment 2: Effect of frequency.....	78
2.5.1 Methods	78
2.5.1.1 Input	78
2.5.1.2 Training.....	78
2.5.1.3 Analysis.....	79
2.5.2 Results and Discussion	79
2.6 Experiment 3: Effect of grammatical complexity	83
2.6.1 Method.....	83
2.6.1.1 Input	83

2.6.1.2 Training.....	84
2.6.1.3 Analysis.....	84
2.6.2 Results and Discussion	85
2.7 General Discussion	86
2.7.1 Word learning mechanisms	87
2.7.2 Role of input.....	90
2.8 Conclusions.....	91
Chapter 3.....	93
Learning to use words	93
3.1 Abstract	93
3.2 Introduction.....	93
3.3 Methods	96
3.3.1 Participants	96
3.3.2 Stimuli	96
High/Known:	98
3.3.3 Neuropsychological measures	98
3.3.4 Procedure.....	98
3.3.5 Recording	99
3.3.6 Data Analysis.....	100
3.4 Results and Discussion	100
3.4.1 Accuracy	100
3.4.2 Context sentence ERPs.....	102
3.4.3 Plausibility ERP effects	103
Chapter 4.....	109
Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context.....	109
4.1 Abstract	109
4.2 Introduction.....	110
4.3 Methods	116
4.3.1 Participants	116
4.3.2 Materials	117

4.3.3 Sentences	117
4.3.4 Word-Pairs	118
4.3.5 Procedure.....	119
4.3.6 Electrophysiological recording.....	121
4.3.7 Data analysis	121
4.4 Results	123
4.4.1 Behavioral performance	123
4.4.2 ERP data: N400 amplitude	126
4.4.2.1 Context sentence endings.....	126
4.4.2.2 Priming task.....	126
4.4.3 ERP Data: N400 distribution.....	132
4.4.3.1 Context Sentences.....	133
4.4.3.2 Priming task.....	134
4.4.4 Relationships between N400 and RT	134
4.5 Discussion.....	136
Chapter 5.....	140
Getting it right: Word learning across the hemispheres.....	140
5.1 Abstract	140
5.2 Introduction.....	141
5.2.1 The role of the cerebral hemispheres in encoding verbal information.....	145
5.2.2 The role of the cerebral hemispheres in the representation of verbal information.....	146
5.3 Method.....	151
5.3.1 Participants	151
5.3.2 Materials	151
5.3.2.1 Sentences	152
5.3.2.2 Word-Pairs.....	153
5.3.3 Procedure.....	154
5.3.4 Electrophysiological recording.....	155
5.3.5 Data analysis	156
5.4 Results	157
5.4.1 Behavioral responses	157

5.4.1.1 Accuracy	157
5.4.1.2 Reaction Time	159
5.4.2 Electrophysiological responses	161
5.4.2.1 N100 amplitude	162
5.4.2.2 N400 amplitude	169
5.4.2.3 N400 distribution	170
5.4.2.4 LPC	173
5.4.2.5 LPC Distribution	174
5.4.3 Correlation between reaction time and N400	177
5.4.4 Correlation between LPC and reaction time	177
5.4.5 Nonword priming	180
5.5 Discussion	182
Chapter 6	189
General Discussion	189
6.1 Lifetime language experience	190
6.1.1 Amount of input	191
6.1.2 Frequency of input	194
6.1.3 Structure of input	195
6.1.4 Conclusions	196
6.2 Immediate sentence context	198
6.2.1 Learning to use words	201
6.2.2 Integrating words into semantic memory	202
6.2.3 Word learning across the hemispheres	204
6.2.4 Conclusions	205
Appendix A	208
Average Precision	208
Appendix B	209
Semantic relationships between categories	209
Appendix C	211
Examples of Sentences used in the Study	211
References	213

LIST OF FIGURES

Figure 1-1. Example of a simple recurrent network.....	25
Figure 1-2. Hidden Unit representation of vocabulary items in a young network before extensive training, after 20,000 sweeps and then at the end of training at 140,000 sweeps...	26
Figure 1-3. Close-up of hidden unit cluster of noun items in older network.....	27
Figure 1-4. Example ERP waveform plotted at a single electrode with a number of common components labeled. Negative is plotted up, and time from stimulus onset is plotted across the X-axis in 100 ms intervals, starting with a 100ms pre-stimulus baseline.	35
Figure 2-1. Example of a simple recurrent network.....	64
Figure 2-2. Hidden Unit representation of vocabulary items in a young network before extensive training, after 20,000 sweeps and then at the end of training at 140,000 sweeps...	65
Figure 2-3. Close-up of hidden unit cluster of noun items in older network.....	66
Figure 2-4. Word learning across size of input. Average node activation is plotted across corpus size for the prediction of the new word in a novel context.	73
Figure 2-5. Change in average precision values over training for each corpus size.....	74
Figure 2-6. Average node activation across average precision value for each corpus size.	76
Figure 2-7. Average precision values over training and frequency condition.....	80
Figure 2-8. Rate of word learning at 80k, 100k and 140k sweeps in both frequency conditions as measured by average word node activations over training.	81
Figure 2-9. Differences between even and uneven new word node activations plotted across difference in average precision values.....	82

Figure 2-10. Average precision values across training in simple and complex input conditions	85
Figure 2-11. Rate of word learning across complexity of input.	86
Figure 3-1. ERPs to Known and Unknown target words in context sentences at midline electrode sites.....	105
Figure 3-2. Plausibility effect for Verbs preceding Known and Unknown targets. Event-related potentials at right-medial central electrode (RMCe) are plotted for plausible and implausible verbs in each condition.	106
Figure 4-1. Diagram of the electrode positions and labels	122
Figure 4-2. Mean reaction time to target items on priming task.....	125
Figure 4-3. Grand average ERPs to known and unknown target words in context sentences at medial electrode sites.	127
Figure 4-4. Grand average ERPs across all subjects and electrode sites for targets in the priming task that were initially preceded by Known and Unknown words that initially appeared in highly constraining sentence contexts.....	128
Figure 4-5. Grand average ERPs across all subjects and electrode sites for targets in the priming task that were initially preceded by Known and Unknown words that initially appeared in weakly constraining sentence contexts	129
Figure 4-6. Grand average ERPs to target words in priming task at the vertex electrode (MiCE).	130
Figure 4-7. N400 Mean amplitude measured from 250-500ms averaged across four electrode sides, MiCE, RMCe, LMCE, MiPA for target words across four prime conditions.....	131

Figure 4-8. Correlations between reaction time and N400 amplitude for all priming conditions.	135
Figure 5-1. Diagram of electrode positions and labels.....	157
Figure 5-2. Mean Accuracy of behavioral responses across experimental conditions.	159
Figure 5-3. Mean reaction time of behavioral responses across experimental conditions.....	160
Figure 5-4. Effect of stimulus lateralization on the ERP. ERPs are plotted across two electrodes where N1 effects of typically largest as a function of visual field of stimulus presentation, collapsed across constraint, prime relationship, and word type. N1 responses (100-200ms) are largest in the contralateral field of stimulus presentation.	163
Figure 5-5. Grand average ERPs across all subjects and electrode sites for targets presented to RVF/LH and LVF/RH that were preceded by Known primes presented in High constraint sentence contexts.	164
Figure 5-6. Grand average ERPs across all subjects and electrode sites for targets presented to RVF/LH and LVF/RH that were preceded by Unknown primes presented in High constraint sentence contexts	165
Figure 5-7. Grand average ERPs across all subjects and electrode sites for targets presented to RVF/LH and LVF/RH that were preceded by Known primes presented in Low constraint sentence contexts	166
Figure 5-8. Grand average ERPs across all subjects and electrode sites for targets presented to RVF/LH and LVF/RH that were preceded by Unknown primes presented in Low constraint sentence contexts	167
Figure 5-9. Grand average ERPs to target words in priming task at a single electrode, MiCE	168

Figure 5-10. Spline interpolated isovoltage maps plot the distribution of the priming effect (Unrelated – Synonym/ID) from 250-500ms across experimental conditions in both visual fields.	172
Figure 5-11. Spline interpolated isovoltage maps plot the distribution of the priming effect (Unrelated – Synonym/ID) from 500-900ms across experimental conditions in both visual fields.	176
Figure 5-12. Correlation of reaction time and N400 across all priming conditions in each hemisphere in A) at a single vertex electrode (MiCe) and B) across all scalp electrodes.....	178
Figure 5-13. Correlation of reaction time and LPC amplitude across all priming conditions in each hemisphere in A) at a single vertex electrode (MiCe) and B) across all scalp electrodes.	179
Figure 5-14. ERP waveforms to Nonwords preceded by Known and Unknown word primes at a single vertex electrode (MiCE).	182

LIST OF TABLES

Table 2-1. Number of word types in each condition.....	71
Table 2-2. Word learning means and standard deviations for different training corpus sizes.	75
Table 2-3. Average precision means and standard deviations for different training corpus sizes	76
Table 3-1. Sentence examples in each condition.....	98
Table 3-2. Accuracy of responses to plausibility task	101
Table 3-3. Correlations between neuropsychological measures and N400 plausibility effects in all experimental conditions.....	104
Table 4-1. Examples of sentences and word pairs in each condition	118
Table 4-2. Mean reaction times (ms) and mean percentage of correct responses for priming task in all conditions.....	125
Table 4-3. F-values from pairwise ANOVAs comparing mean amplitude N400 to related, unrelated, and synonym/ID targets in four main prime conditions	132
Table 5-1. Examples of sentences and word pairs in each condition	152

ACKNOWLEDGMENTS

This dissertation could not have been completed without the support, advice, and encouragement of my advisors, colleagues, labmates, friends and family.

I have been very lucky to learn from mentors who are not only great scientists, but are warm and generous people. I must first thank Jeff Elman, for being my advisor, advocate and all-around nice guy. He has given me invaluable advice about the navigating scientific and nonscientific aspects of academia and allowed me much freedom and flexibility in this project, but has always been close at hand when I needed help. Marta Kutas welcomed me into her lab and provided much helpful feedback about this project along the way. I am especially grateful to Nina Dronkers for her kindness and encouragement over the years.

My life has been enriched by the many wonderful and intelligent people who I have known in the UCSD Cognitive Science community, the Center for Research in Language and Kutaslab. Thanks to Jenny Staab for interesting discussions and board games, Michael Kiang, Katherine DeLong and Kim Sweeney for helping to point me in the right directions in my early days in Kutaslab. I am also thankful to my fellow graduate students for their unique intellectual perspectives, their camaraderie and commiseration.

Bob Buffington, Luis Palacios, Esmeralda de Ochoa and Paul Krewski provided indispensable technical help. I am especially grateful to Bob fixing my lab account after I've made a mess of things (multiple times), especially at odd hours.

I am deeply thankful to my friends and family for their support. I am deeply indebted to James Cavanaugh, who convinced me to stay in graduate school after the death of my advisor and who reminded me about the joys of research during even those difficult times. I

am also grateful to Matthew Stumpf and Anne Furgerson, who cheered me on as I wrote, and, whenever I needed a break, fed me fabulous food, and lent me a bit of earth to grow my “dissertation squash”. Most importantly, I want to thank my parents, especially my father, who fostered my scientific curiosity from a young age, always did their best to encourage me to think independently and helped to give me the tools to answer my (seemingly endless) “whys”.

Lastly, I want to thank Liz Bates, who initially encouraged me to come to UCSD. Although the time in which I knew her was too brief, her unmatched enthusiasm for science, thoughts and writings have had a significant influence on my own work, and continue to be a rich source of inspiration.

Chapter 2, in full, is a reprint of the material as it appears in Borovsky, Arielle & Elman, Jeff. (2006). Language input and semantic categories: A relation between cognition and early word learning. *Journal of Child Language*, 33, 759-90. The dissertation author was the primary investigator and author of this paper. Permission to reprint was granted by Cambridge Journals, Cambridge, UK.

Chapter 3, in part, has been submitted as it may appear for publication; Borovsky, Arielle, Elman, Jeff, and Kutas, Marta. (2008). *Learning to use words: Event related brain potentials index single shot word learning from context*. Under review in *Cognition*. The dissertation author was the primary investigator and author of this paper.

Chapters 4, and 5 appear in preparation for publication by Borovsky, Arielle, Elman, Jeff and Kutas, Marta. The dissertation author was the primary investigator and author of these papers.

VITA

- 2001 B.S. in Psychology, University of Florida
- 2001-2002 Research Assistant, NIMH Center for Study of Emotion and Attention
- 2004 M.S. in Cognitive Science, University of California, San Diego
- 2003-2008 NSF Fellow, NIH Training Grant Fellow, Teaching Assistant, University of California, San Diego
- 2004-2006 Editor, Cognitive Science Online
- 2007-2008 Editor, CRL Newsletter
- 2008 Ph.D. in Cognitive Science, University of California, San Diego

Publications

- Borovsky, A., Elman, J & Kutas, M. (2008). *Learning to use words: Event related brain potentials index single shot word learning from context*. Under review in *Cognition*
- Borovsky, A., Elman, J. & Kutas, M. (2007). Getting the gist is not enough: An ERP investigation of contextual word learning. *Proceedings of the 29th Annual Cognitive Science Society* (pp. 119-124). Austin, TX: Cognitive Science Society.
- Borovsky, A., Saygin, A. P., Bates, E., Dronkers, N. F. (2007). Lesion correlates of conversational speech deficits. *Neuropsychologia*, 45(11), 2525-2533.
- Borovsky, A. Elman, J. (2006). Language input and categories: A relation between cognition and early word learning. *Journal of Child Language*, 33(3), 759-790.

ABSTRACT OF THE DISSERTATION

Word learning in context:

The role of lifetime language input and sentential context

by

Arielle Borovsky

Doctor of Philosophy

University of California, San Diego, 2008

Professor Jeff Elman, Co-Chair

Professor Marta Kutas, Co-Chair

Our experience with words defines how we understand them. In this dissertation, I examine how two kinds of experience influence how words are learned - that of “global” lifetime language experience, and “local” experience from immediate linguistic context. Computational simulations are used in the first set of experiments to simulate a variety of early language learning environments that vary in amount, frequency, complexity of linguistic

input. Brainwave experiments are used in the second set of studies to probe the neural and cognitive correlates of word learning from sentential context.

The first set of studies are computational simulations that explore how differences in linguistic experience can explain differences in word learning ability and organization of semantic knowledge. We varied the amount of language input, sentential complexity, and the frequency distribution of words within categories. In each simulation, improvements in category structure were tightly correlated with subsequent improvements in word learning ability. These simulations suggest that vast differences in lexical proficiency in children can at least partly be explained by differences in early language environments and underlying cognitive abilities like categorization.

The second set of studies explore how local experience influences single-shot word learning in a series of three brainwave studies. Adult participants read known and unknown words in high and low constraint sentences and then made plausibility judgments on their usage in subsequent sentences, or saw these words again as primes in a semantic priming task. These studies found that participants were able to integrate knowledge about the meaning and usage of unknown words that initially appeared in highly constraining contexts, but not low constraint contexts. In addition, a lateralized version of the semantic priming probe task revealed that the right hemisphere initially participates in the semantic priming of novel words. Together these studies highlight the importance of experience in acquisition of word meaning, and reveal that the brain is able to quickly acquire significant and sophisticated information about word meanings after only a single instance. In addition, these studies suggest a method by which lexical acquisition can be measured via electrical brain activity.

CHAPTER 1

BACKGROUND AND SIGNIFICANCE

"One cannot guess how a word functions. One has to look at its use and learn from that." (Wittgenstein, 1953)

Humans have a unique and universal capacity to learn many thousands of words- an ability unmatched by any other species. How is it that we are able to acquire so many words? As Wittgenstein (1953) suggested, much of word learning occurs via language use itself. That is, words are learned due to our experience with them as they are used in the language. As we hear an unfamiliar word, we take cues from the context in which it appears in order to understand its meaning. This context can be a number of things, including the immediate sentence context in which a novel word appears, the greater discourse context, the situational constraints, world knowledge, and knowledge from previous language experience. In this dissertation, I examine how two aspects of context influence lexical acquisition. The first, examined in Chapter 2, is what I term "global experience" or, experience gained from hearing a language over a lifetime. The second is examined in Chapters 3, 4, and 5 and occurs through the immediate sentential context that surrounds a novel word, which I call "local experience."

Given that multiple kinds of language experience are highly important in word learning, it is surprising that the majority of word learning research has not examined these factors. An enormous body of research on word learning in young children and adult bilingualism has accumulated over the past fifty years. Yet, the primary focus of this research

has largely been on the strategies and factors involved in the explicit learning of object names (see Bloom, 2000; Golinkoff et al., 2000 for reviews). Common questions in this research include: What words make up the lexicon across development (i.e., Brown, 1973; Nelson, 1973)? How does one infer the proper referent for a word (Golinkoff, Mervis, & Hirsh-Pasek, 1994)? When is the concept of noun or verb understood (i.e. Olguin & Tomasello, 1993; Tomasello & Olguin, 1993)? What is the developmental trajectory of word learning (Bates & Goodman, 1997)? What is the relationship between the development of word learning and other cognitive abilities (e.g. Gopnik, Choi, & Baumberger, 1996)? While early word learning studies were concentrated upon gross surveys and interviews of adult and child lexicons (Bloom, 1970; Braine, 1963; Brown, 1973; Nelson, 1973), presently, many of these questions are often explored in experimental paradigms that probe explicit word learning of a new object or set of objects. For instance, one oft-cited paradigm tests how explicitly defined object names are generalized to other similar objects based on shape, size, color, etc (P. Bloom, 2000). Another common test probes how social cues such as gaze and pointing may lead to novel word meaning acquisition for objects (Baldwin, 1991). This work has yielded many valuable insights into the mechanisms involved in word learning. However, a basic limitation of this work is that the majority of words that we know are not objects and their meanings are not learned explicitly. Rather, the majority of novel word meanings are acquired incidentally, via linguistic context (Nagy, Anderson, & Herman, 1987; Sternberg, 1987). For instance, it is estimated that when a child enters school, he or she will learn on average 1200 new word families per year, mostly through reading (Anglin, 1993).

Despite the fact that a large number of words are learned implicitly via context, there is huge variability across individuals in ability to learn from context, and consequently, large individual differences in vocabulary level. For instance, estimates of the number of words that adults come to know range between 40000-150000 words (Aitchinson, 1994; Beck & McKeown, 1991; Bloom, 2000; Pinker, 1994). It is clear from these estimates that adults and children learn an enormous number of words – but that the range between individuals can be huge. What factors account for this variation in lexical knowledge? It is arguable that at least some of this variation can be accounted for by differences in linguistic experience. Further, when measuring the likelihood of acquiring a single word, there are a number of experiential factors that could play a role. For one, the previous experience of the learner could be important. If an individual already knows many words, it may be possible that acquiring additional lexical items could be easier - especially for words that are similar in meaning to which the learner is already familiar. On the other hand, it is also possible that the immediate context in which a word appears could be more or less helpful in acquiring the word's meaning. The main goal of this dissertation is to examine how these two kinds of experience – “global” and “local” - influence the cognitive and neural correlates of contextual word learning.

In order to better understand the state of knowledge regarding both kinds of experience on word learning, I review the relationship of global and local experience to lexical acquisition below. First, I examine how global experience can influence early lexical development and describe a number of theories that have attempted to explain developmental improvements in vocabulary acquisition. Following this, I discuss the role of

more immediate contextual factors on word learning, and examine the neural and cognitive correlates of rapid contextual word learning.

1.1 The effects of lifetime language experience on early lexical and cognitive development

One of the most well studied aspects of word-learning is termed the “vocabulary spurt” This phenomenon occurs around the age of 18 months or around the time 50 words are known, when children begin to acquire new words quickly. Typically, first words appear around the age of 12 months, but are learned very slowly up until the vocabulary spurt at 18 months. At this point, rate of word learning accelerates noticeably such that by 24 months some children may comprehend up to 900 words (Carey & Bartlett, 1978).

While the vocabulary spurt is exhibited by most toddlers, there is also tremendous variation in the rates at which early words are learned and in how many words come to be known (Dale & Fenson, 1996). Traditionally, there has been an implicit assumption in educational systems that vocabulary level can be linked to intelligence or a general capacity to learn words. However, there is some evidence that these vocabulary differences arise due to large differences in early language exposure (Hart & Risley, 1995; Huttenlocher, Haight, Bryk, & Seltzer, 1991; Snow, Burns, & Griffin, 1998). For example, Hart and Risley (1995) recorded and transcribed child-directed speech (CDS) in the homes of 42 children once a month between the ages of about ten months to three years. These children were classified into three groups based on socioeconomic status (SES): professional, working, and welfare. There were large differences between SES groups in amount of language these children heard. Based upon the quantity of CDS recorded at these monthly visits, they extrapolated

that children in the professional group could expect to hear 11 million words by the age of three, whereas those in the welfare group would hear only 3 million words. These differences in input were significantly related to vocabulary level of the children over this period. By the age of three, children in the professional group attained word comprehension levels of approximately 1100 words, while the welfare group knew only about 500 words.

At the same time, there are robust and numerous findings that suggest vocabulary size is a good predictor of reading comprehension ability (i.e. Anderson & Freebody, 1981; Cunningham & Stanovich, 1997; Stahl & Fairbanks, 2006) and ability to learn new words in context (Cain, Oakhill, & Lemmon, 2004; Elshout-Mohr & van Daalen-Kaptein, 1987; McKeown, 1985). For instance, Cunningham and Stanovich (1997) report that vocabulary scores in first grade predicts a significant amount (30%) of variability in reading comprehension ten years later, in 11th grade. Further, early linguistic experience at home is associated with a child's ability to learn from linguistic input at school (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002). For instance, a child who has had the benefit of an enriched early language environment can also be expected to retain this advantage in school by learning new words more quickly. Altogether it is quite likely that impoverished language exposure in infancy can lead to school-aged vocabulary and reading comprehension deficits. However, while it may be difficult to make up for "lost experience" in overall quantity of input – it may be possible that not all language experiences are equally valuable.

In fact, some evidence indicates that the quality or structure of input may influence vocabulary growth. For instance, it is possible to imagine a situation where language exposure might be overly repetitive and not contain much useful information, or it might be too complex for a child to understand or parse. Studies of child-directed speech (CDS)

(Snow & Ferguson, 1977) have indicated that the language that children hear is simpler in both syntactic structure and the kinds of words used. One ambitious study of CDS to 12 children between the ages of 2 and 3 years old (Cameron-Faulkner, Lieven, & Tomasello, 2003) found that 20% of all the sentences heard in this period are sentence fragments, which usually are responses to a question, and that another 32% are questions. Complex sentences only accounted for 6% of all utterances heard by children at this age. These results were also strikingly similar to a similar, but smaller study by Wells (1981).

It is possible that simpler structure of CDS could improve word learning by reducing the processing demands placed on the child when a new word is encountered. In order to support this idea it is necessary to compare the kinds of input heard with the words actually learned by the child. To this end, Brent and Siskind (2001) examined the relationship between single word input to infants aged 9-15 months with vocabulary level at 18 months – i.e., at the beginning of the vocabulary spurt. By examining this relationship, the investigators were able to study how the words heard in the simplest kind of grammatical construction affect the learning of that word. They found a reliable association between number of times word is heard in isolation and comprehension of that word at 18 months. This association was stronger than solely measuring the total number of times the word was heard in utterances of varying complexity. When taken into account with the other CDS research described above, this work indicates that word learning as whole might be easier when sentence structures remain simple.

On the other hand, there are also recent findings that indicate increased syntactic complexity in CDS leads to improved vocabulary at age 2 (Hoff & Naigles, 2002). However, here syntactic complexity was not measured through detailed constructional analysis, but

through mother mean length of utterance (MLU). Hoff and Naigles (2002) concluded that increased syntactic complexity, not less, is important in aiding word learning.

Since advantages from enriched language exposure do persist across development (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002), it may be that input could change overall word learning proficiency. That is, it might be that there are fundamental changes in the learning mechanisms by which words are learned that are altered solely by exposure to language. It is therefore a major goal of the first set of studies in this dissertation to examine how the quantity and quality of language input can change how new words are learned.

Concurrently, while it is clear that the role of language input is important in shaping expected language proficiency, this does not provide a straightforward explanation as to why word learning rates change around the time of the vocabulary spurt or why better rates of lexical acquisition are maintained in children who receive better input. Exploring at least one explanation of this is another main goal of this work. However, this also leaves open the question as to what kinds of learning mechanisms may be responsible for these input driven effects on lexical learning. In order to narrow the range of possible mechanisms, the next section considers some past and present theories regarding the onset of the vocabulary spurt.

1.1.1 Vocabulary spurt theories

Much work has attempted to explain rapid word learning observed in children and adults by proposing mechanisms that change around the onset of the vocabulary spurt. Five main mechanisms have been posited to explain why the vocabulary spurt occurs:

1) Naming insight: It is the result of a “naming insight”, where the child realizes that all objects can be named (McShane, 1979).

2) Social/pragmatic development: The child has developed a social/pragmatic expertise in order to discern a speaker’s communicative intent (Baldwin, 1991; Tomasello, 1996).

3) Constraints and Principles: Sophisticated word learning constraints guide the child to make good “guesses” at what words may mean. (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Woodward & Markman, 1998)

4) Connectionism: Non-linear patterns emerge from gradual learning by association (Marchman & Bates, 1994; Plunkett, Sinha, Moller, & Strandsby, 1992)

5) Categorization: There is an overall improvement in other, non-linguistic abilities, such as categorization (Gopnik & Meltzoff, 1987, 1992).

Below, each is reviewed in detail.

1.1.1.1 Naming insight

This is one of the earliest proposals about what kind of learning mechanisms are responsible for the vocabulary spurt. The essential idea behind the naming insight is that some kind of switch “turns on” in the child’s head with the realization that words have symbolic referents (McShane, 1979). It is proposed that this insight should correlate with the vocabulary spurt, and that word learning should increase because of it. However, there has been little follow-up work with this idea, most likely because it is not clear how to actually test this hypothesis.

1.1.1.2 Social/pragmatic development

In this view, children begin to understand that there is a communicative intention underlying adult language and that word learning is aided by discerning these intentions. Observations that infants can follow a speaker's gaze to learn a new word, rather than attach a word label to an object they are attending to at that moment (e.g., (Baldwin, 1991; Tomasello, 1996) is cited as evidence in support of this view. Additionally, this ability appears around the time of the vocabulary spurt. Before this period, children tend to rely on joint attention between themselves and the speaker to establish labels for objects. Because this reduced reliance on mutual engagement to properly establish referents for new words appears around the time children are expanding their lexicon very rapidly, supporters of this social cognitive approach have proposed that it is these abilities that enable children to quickly learn words. However, proponents of these theories do not maintain that social cues are the only word learning mechanisms, which leaves open discussion about what other influences might be responsible for early word learning.

1.1.1.3 Constraints & Principles

The next group of theories arises from a tradition where language is considered a "special" ability that is innate to humans and that it is not reliant upon other kinds of cognitive abilities. Therefore, theories that fall under this section, attempt to seek out principles that are specialized for word learning only. Two major theories have been offered on this account. The first of these are called "constraint theories" (Markman, 1989; Woodward & Markman, 1998). The idea is that word learning is guided by a set of innate constraints that bias a child to interpret new referents as a type of label. The first of

these proposed constraints was the whole-object assumption, which was put forward as a way to explain findings from several studies (see Markman, 1989) that found that children were more likely to assume that a new word referred to a whole object, rather than an interesting or salient part of it. Additional constraints in this same vein were proposed, the most widely known are: the mutual exclusivity assumption (Markman, 1989; Markman, Wasow, & Hansen, 2003), which hypothesizes that children will only apply a new word to an object they don't already know the name for. Another constraint, the taxonomic bias, proposes that new words are assumed to refer to basic-level categories. For example, the word 'dog' would be assumed to refer to most hairy animals with four legs that bark, but not just "golden retrievers" or "all mammals". While the constraints described here are the most commonly studied, there are also a number of others that are less prevalent.

This proliferation of constraints prompted Golinkoff, Mervis and Hersh-Pasek (1994) to simplify this approach through their "lexical principles" model. This framework consisted of six principles organized into two tiers. It was proposed that the three principles on the first tier operate initially on word learning, and that these are domain general principles that do not need to apply only to language. However, the second tier, which includes the conventionality, categorical scope, and novel name-nameless category (N3C) principles, applies only to word learning. Many of the constraints that were proposed through earlier work were accounted for in the second domain specific tier of this framework, and provided an explanation for why patterns of word learning change over development. Also, the usage of principles from the second tier is considered to mark the beginning of the vocabulary spurt.

However, recent empirical and theoretical criticisms of this model have arisen. Notably, findings that principles of extension and mutual exclusivity do not apply just only to word learning, but facts (Markson & Bloom, 1997) and action learning (Childers & Tomasello, 2003) have led some researchers to suggest that previous lexical-specific principles might instead be a product of domain general learning mechanisms that apply to all types of learning. Constraint theorists have replied with findings showing that toddler's extension of new words and facts do not show equal levels of performance (Behrend, Scofield, & Kleinknecht, 2001; Waxman & Booth, 2000). However, Childers and Tomasello (2003) argued that extensions of facts are a poor mode of comparison with extension of words, because they have different logical entailments. Bloom and Markson (2001) point out that knowing how to extend the name to of a recently learned objects such as "koba", does not necessarily extend in the same way that a fact such as "this is the one my uncle gave me." Additionally, they point out that facts are able to refer to ideas that are not easily encapsulated by words such as, "this is the sort of desk they used to sell at Pottery Barn". Instead of comparing facts with words then Childers and Tomasello (2003) showed that extension of a non-verbal action that can be described as, "this is the sort of action that we do with things like this" is similar to the kinds of extension seen in new words.

1.1.1.4 Connectionism

An alternative to the constraints and principles view is one that words are learned through domain-general processes that make note of statistical associations between words and other properties that accompany their referents. These theories also try to account for patterns pointed to by the constraints theories as resulting from general

attentional biases that make certain properties more perceptually salient, such as object shape (Smith, Jones, & Landau, 1996). Work under this account has focused on two separate areas.

The first of these explored how children come to show certain constraint-like behavior when determining the referent for a new word. Explanations for constraints are then recast in terms of frequency of occurrence and object saliency. Essentially, if a particular object or salient feature, has been paired with a word enough times then the child has no choice but to learn that this word must refer to the common denominator in all these experiences. This “dumb attentional mechanism” then is described as an unconscious and automatic pattern associator that is influenced both by cues that are present at the moment a word is heard, as well as by cues that have mattered previously. This allows for both perceptual and conceptual knowledge to be utilized when learning a new word.

Aside from being able to explain much of the constraint based literature, there is some empirical support for the idea that children are using automatic processing for new word learning. Studies that have observed the types of object name generalizations that 3 year olds make have shown that children tend to exclude some information in preference of others. For instance, (Smith, Jones, & Landau, 1996) report on several studies which demonstrated that children tend to extend the name of an object to another object based on similar features, but not to those with similar functions, even though children did exhibit knowledge in a non-naming task of the functional similarity between these objects. The authors take this result to suggest that the act of naming itself forces children to

attend to particular cues that have proven themselves as reliable best guesses to referents of new words.

The second type of theory has examined how reliable patterns in language provide information that aid in word learning. The role of linguistic input here is seen as a means to constrain the problem when a new word approaches by paring down the likely candidates through information in other parts of the sentence. Gillette, Gleitman, Gleitman and Lederer (1999) demonstrated with human word-learning simulations on adults that information such as noun co-occurrence and structural cues can indeed be useful in learning new verbs.¹

Although the two areas of research outlined in this section might not seem compatible, at their theoretical core they are quite similar. It is true that each of these theories point to a different kind of input that come into use for learning words. The first emphasizes perceptual cues that may occur often and are salient to the child. The second proposes that statistical regularities in linguistic input form reliable patterns that constrain the problem space. However, the underlying mechanism is identical in both cases. Both accounts propose that a connection between the input and new word is formed through a statistical association that is guided by domain general learning principles that do not apply only to words.

¹ Although Gillette et. al. (1999) do not interpret these findings as consistent with a connectionist account. However, it seems that the way in which noun co-occurrence information aids in new verb learning in this study could be consistent with an associative learning mechanism.

1.1.1.5 Categorization

This account of word learning proposes that the realization that words can be classified into categories improves the rate at which new words can be learned. While it is difficult to probe young children's knowledge of semantic structure, it is often indirectly measured by various categorization tasks. For example, exhaustive sorting tasks are a popular kind of paradigm that measures the toddler's ability to place a number of objects in categories (Gopnik & Meltzoff, 1987, 1992; Mervis & Bertrand, 1994, 1995; Poulin-Dubois, Graham, & Sippola, 1995). This line of research has shown that the categorization abilities tend to improve around the time of the naming explosion. Some criticisms of these findings arise from observations that some children who have obtained vocabularies much larger than the average 50 words did not seem to undergo a vocabulary spurt (Goldfield & Reznick, 1990, 1996). However, these studies did not measure categorization abilities. Mervis and Bertrand (1995) followed several children from an earlier study who had gained a sizeable vocabulary (between 63-123 words) but who had not yet exhibited a spurt. They reported that these children exhibited a late spurt, and that it coincided with the onset of exhaustive sorting abilities. Although these later papers debated the existence of the vocabulary spurt, the careful reconsideration of this phenomenon with additional measurement of categorization ability provides strong evidence that changes in conceptual structure do occur along with word learning. Yet, since the ability to test and retest toddlers is necessarily limited, the relationship between categorical development and vocabulary spurt has only been examined at an indirect and somewhat crude level. Therefore, one goal of the present work is to re-examine how changes in category structure and categorization of new words might affect new word learning.

As different as all of these theories may seem, they are not mutually exclusive. There is an increasing sentiment amongst developmental scientists that a variety of strategies and learning mechanisms can aid word learning. In support of this view, some theoretical proposals (i.e., Bloom, 2000; Golinkoff et al., 2000) have attempted to merge these ideas. Additionally, recent proposals have criticized previous trends that pit theories against each other in order to highlight dichotomies. Instead, it has been suggested that holistic accounts that can incorporate multiple factors are preferable (Bloom, 2000).

To review, the aim of the first set of studies in this dissertation is to examine the relationship between global linguistic experience and lexical and categorical development. We propose that associations from linguistic input itself can change underlying cognitive mechanisms like categorization, and that development of categorical structure can in turn boost in word-learning. Connectionist accounts are the most compatible with this type of input-driven learning. Since these studies examine the relationship between categorical and lexical learning, they are also highly relevant to the categorization account of word learning. In addition to the categorization theory of the vocabulary spurt, there are more general debates regarding the relationship between cognition and language in the developmental literature. In the next section, I provide a general review of the various positions in this debate.

1.1.2 Issues in cognitive and linguistic change

Word-learning scientists have begun to integrate multiple theoretical standpoints into their research commentary by acknowledging that there may be multiple factors that may account for word learning. However, the relationship between cognition and language

has not undergone this kind of renaissance. In this field, there is still debate about whether cognitive and linguistic learning belong to separate systems, and whether the changes that appear in both around the time of the vocabulary spurt are related to a gradual or abrupt change. Both of these issues are considered in turn.

1.1.2.1 Continuous vs. Discontinuous

At 18 months of age, there are a number of cognitive changes that occur in the child in many domains, not just in language and categorization. Piaget proposed that around the end of the second year these advancements mark a new stage of development where fundamentally different cognitive mechanisms become available to the child. In this view, development occurs in a series of discrete, discontinuous stages that must occur in a particular order, with each stage being more advanced than the last. However, this has been a point of some contention because more recent work has revealed that many of the developmental milestones that were observed by Piaget could be detected at earlier ages when certain task demands were reduced.

As for categorization, the ability to sort objects exhaustively appears around 18 months, but it is possible to observe categorization abilities in younger infants as well. For instance, in tasks where the motor and linguistic demands of the categorization task have been reduced, infants as young as 9 months old will sequentially touch all objects in an array that belong to one category (Sugarman, 1981). However, there is much disagreement about how performance on sequential touching tasks relates to the later categorization abilities (Bauer, Dow, & Hertsgaard, 1995; Thomas & Dahlin, 2000). Most recently, sequential touching studies tend to ask questions that have more to do with what

order categorization abilities appear in children, rather than trying to pin-point “the age” at which categorization is possible. This general trend suggests that categorization abilities emerge gradually over a continuum, rather than due to a fundamental change at 18 months.

Similarly, there has been deliberation about whether lexical acquisition around 18 months can be explained by single or separate learning mechanisms (Bates & Carnevale, 1993; Golinkoff, Mervis, & Hirsh-Pasek, 1994; Nazzi & Bertoncini, 2003; Plunkett, Sinha, Moller, & Strandsby, 1992). The shift from fast to slow word learning observed around the vocabulary spurt has often been interpreted as a change in underlying strategy or cognitive machinery. However, findings by Plunkett et al (1992) and Bates and Carnevale (1993) suggest that non-linear changes in growth curves of learning can be accounted for by gradual changes of a single mechanism. For instance, Plunkett et al (1992) trained a neural network to associate images and object labels. Without making any functional or structural changes to the network, it exhibited a kind of “vocabulary spurt” where labels and images pairs were learned with increased speed. The central point is that the network exhibited dramatic, sudden, non-linear changes in its performance that are similar to the patterns one might see in children during the vocabulary spurt, yet no there were no modifications in either the input or architecture of the network across time. While probing the nature of lexical learning mechanisms in toddlers poses a difficult task, these simulations suggest that abrupt changes in the rate of learning do not necessarily relate to discontinuous changes in the learning mechanism. However, understanding precisely what drives such changes remains to be understood.

1.1.2.2 Domain-general vs. Domain Specific

There are a number of positions that outline the relationship between cognitive and linguistic development, ranging from standpoints that posit they are completely unrelated, to those that are entirely related. The former extreme maintains that linguistic and cognitive abilities are completely modular or domain specific, so they develop independently of each other. According to this account, semantic structure is not acquired due to language input (Fodor, 1983). Rather, conceptual structure is innate, and words are learned that correspond to this innate “mentalese.” The idea that language is learned by universal linguistic principles (Chomsky, 1980; Pinker, 1991) and does not interact with other aspects of cognition is compatible with this view.

Generally, this view has fallen out of favor with developmental psychologists largely due to overwhelming evidence that there is at least some association between improvement in cognitive and linguistic functioning in infants. The question then becomes to what degree language and cognition are related, and how each may inform the other. These ideas range on a spectrum, from something like the strong version of the Whorfian hypothesis (1956) that states that all thought is impossible without language, to more interactive versions where language and cognition are completely integrated and reliant upon each other.

One of the less interactive positions is inspired by Piaget’s (1971) view of development. This theory proposes that cognitive and linguistic achievements occur sequentially or simultaneously as part of advancement through particular stages. The large number of both cognitive and linguistic changes that occur in toddlers around the end of second year like the appearance of the vocabulary spurt, improved categorization,

deferred imitation, and self-recognition, is cited as evidence for this view (Courage & Howe, 2002). However, Piaget did believe that conceptual development preceded linguistic development. In this view, language developed in order to express new concepts. Consequently, this view proposes that conceptual development drives linguistic development, although there are limits to the expertise which may be attained in both of these domains related to developmental stage.

Other proposals maintain that word-learning might specifically drive conceptual organization (Bowerman, 1996; Choi & Bowerman, 1991). Bowerman cites evidence from cross-linguistic studies that suggests children will readily learn the spatial categorization scheme present in their language. For example, Korean focuses on the fit between objects, while English tends to emphasize the kind of containment or method of support. However, Choi and Bowerman (1991) show that children learning each language have no problem describing these spatial relations in terms appropriate to their language. So, Korean children will describe a peg in a hole as tight or loose fitting, while English-speaking children will say that the peg is “in” or “on” the hole.

Mandler (1996) replies to this claim by proposing that younger children might be able to carve-up space in ways that are not linguistically determined. She suggests that there are a number of different ways that the world can be categorized – all of which are initially available, but that language constrains these possibilities into a regularized convention that differs from language to language. Further, McDonough, Choi and Mandler (2003) find that 9, 11 and 14 month olds all are able to distinguish between tight/loose and in/out distinctions, regardless of linguistic environment.

Here, two main views that are delineated. The first proposes that language is driving the formation of categorical structure (Bowerman, 1996). The second maintains that non-linguistic categories are already formed via the infant's physical experience with the world, but that language constrains the possibilities (see Casasola, 2008 for a review).

Gopnik and Meltzoff's (1992; 1997) theory suggests that cognitive and linguistic development are related to each other, and that both can influence each other equally. For instance, they propose that exposure to a particular word will lead to a drive to learn the underlying concept, but that this new knowledge will then interact with knowledge already in place, which might then lead to new word learning. Evidence for this idea is cited in cross-linguistic studies, which find that the appearance of categorization abilities arise later in Korean speaking children than in English speaking children (Gopnik, Choi, & Baumberger, 1996).

However, it is not unreasonable to say that language can be used as a vehicle whereby one can learn new concepts. Indeed, language conveys semantic content that surely must be of some use to a young language-learner. The thing that still is left to be determined is how the child might use language. Does it aid in carving up categories that are already known? What kind of information does the structure of language provide, and how might this influence category development?

In the first set of experiments in this dissertation, these questions are explored by varying several key factors of linguistic input to neural networks: 1) the quantity of words exposed 2) the frequency of words, and 3) the structural complexity of input. It is therefore possible to examine how these factors may influence categorical formation and subsequent word learning. The use of computational simulations affords the possibility to

explore how this information is learned via linguistic input alone without the influence of additional experience from other perceptual, non-linguistic sources. In addition, simulations make it possible to precisely measure categorical development, which is very difficult to do with children. It is therefore possible to use simulations to examine the nature of the relationship between categorical development and lexical learning. First, we can examine how linguistic input drives categorical development; then, through measuring the speed at which new words are learned after initial training, we can investigate the relationship between categorical development and lexical acquisition.

The use of computer simulations presents unique opportunities to examine theories in ways that may not be as easily done in children. In the next section, we consider some of the advantages of this methodology.

1.1.3 Why simulations?

There are several important reasons why computer simulations are the method of choice for these studies. First, since this research investigates the role of linguistic input upon word learning and category development, it is much easier to systematically alter language exposure to computer network in ways that are experimentally interesting rather than relying upon chance variation that occurs in natural input to children. Connectionist networks are able to generalize to new input, in this case words, on the basis on experiences they have seen before. This type of learning closely coincides with associationist theories of word learning. As a result, computational simulations are an optimal method for examining the kinds of word learning that result from associations and generalization of language without pre-specified biases or constraints. Additionally, neural

networks can be “dissected” and examined in ways that kids cannot. For instance, research in children exploring category development relies mainly upon indirect evidence measured via behavioral categorization abilities. On the other hand, the structure of categorical representations can be directly measured in neural networks. Lastly, running computational simulations is less costly both in terms of time and resources required than in studies with live children. This allows scientists to use computational simulations to specific questions that would be more fruitful and efficient to examine in live children than what might have been possible without computational simulations. Overall, the use of computational simulations can be seen as an aid, but not a replacement for developmental research on children and can reduce the time and resources required to answer important questions that might not otherwise have been possible without their use. In order to better understand how simulations will be useful for our first set of studies, the next section describes how it is possible to use simulations to measure categorical development.

1.1.4 Measurement of category development

As described above, it has been hypothesized that the development of categorical structure of concrete noun categories are related to the surge in productive vocabulary (Gopnik & Meltzoff, 1987, 1992, 1997; Mervis & Bertrand, 1994, 1995). Studies of semantic expertise in children often rely on indirect evidence, such as observations of the ability to exhaustively categorize objects, or to place new objects in learned categories. However, in computational simulations it is possible to explore categorical representations

both by manipulating how well category members fit together, but more uniquely, by measuring the actual representations that form.

Previous simulations by Rogers and McClelland (2006) use the first approach. Here, they examine changes in category coherence of objects that are known to a neural network by manipulating the feature correlation of category members. Category coherence is a term borrowed from Murphy & Medin (1985) that refers to how well the features of a category “fit together”. Previous behavioral work involving category coherence suggests that learning of new category members is indeed improved when coherence is high in regards to what is already known about the category (Rehder & Ross, 2001; (Rehder & Ross, 2001; Wattenmaker, Dewey, Murphy, & Medin, 1986; Wisniewski, 1995).

However Rogers and McClelland (2006) have little to say about how language itself might aid category learning. However, these simulations do fall in line with behavioral observations that how well a new object is learned depends on how similar it is to other category members, as well as how coherent the category is to begin with. However, these simulations were exploring object learning through feature activation, but not through language input. Basically, they worked by presenting inputs to the network that correspond to an object. Training then proceeded by correcting the network’s predictions on the types of features it expected each object to have.

On the other hand, in our approach, the use of language like input allows examination of the kinds of representations that might form when words in a category are related to its occurrence in similar contexts to other words in the category. This will

involve active probing of the network to measure the coherence of representations that the networks have developed from this kind of input.

It should be noted that this kind of exploration does not measure semantic knowledge in the most traditional sense. Conventionally, studies that examine conceptual development tend to emphasize the importance of item-specific features and their relationship to other items in the same category. This type of research is most compatible with the type of simulation conducted by Rogers and McClelland (2006) that was described above. Instead, in the current work ‘semantic’ knowledge is applied more loosely in this context and is intended to explore the kinds of representations that can be formed from linguistic context alone. This is not to say that the influence of non-linguistic information is not also important. However, since the goal of this work is to examine how linguistic input can influence lexical and categorical development, non-linguistic issues are not addressed here.

Measurement of category formation in neural networks is accomplished by quantifying network’s internal (hidden unit) representations of words in a particular category (see Figure 1-1 for an example network). Here, if the network has learned that a set of words, or items within a category are similar in meaning/usage, these items will exhibit similar hidden unit activation values. It is possible to cluster these representations graphically (Figure 1-2, Figure 1-3) to examine the degree of similarity in hidden unit values between each word. Figure 1-2 compares the hidden unit representations of network before and after training in this study. From this example, it is clear that the network has learned not only the differences between nouns and verbs, but also subcategories between them (Figure 1-3). By comparing how close all members of a category are to each

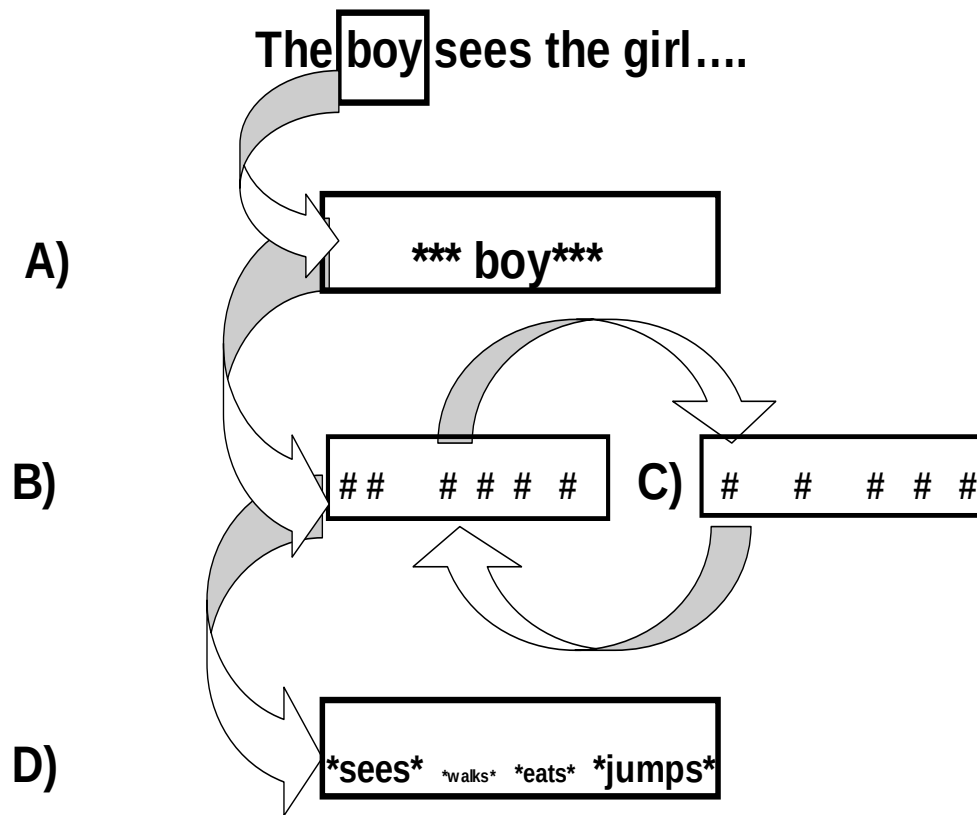


Figure 1-1. Above, words from the sentence “the boy sees the girl” are presented one word at a time to the network to the input layer (A). Next, the word is fed into the hidden layer (B), which also receives information about the immediately previous network states from the recurrent layer, (C). As training proceeds, the hidden layer will develop numerical internal representations that can be used to generalize to similar inputs. In the output layer (D) the network predicts the next possible words after ‘boy’. Generally, the SRNs will learn to predict a range of words that are possible that correspond to their frequency with which they are associated. Here, the network is predicting both the actual next word “sees” but also other possible words like “jumps” and “eats”.

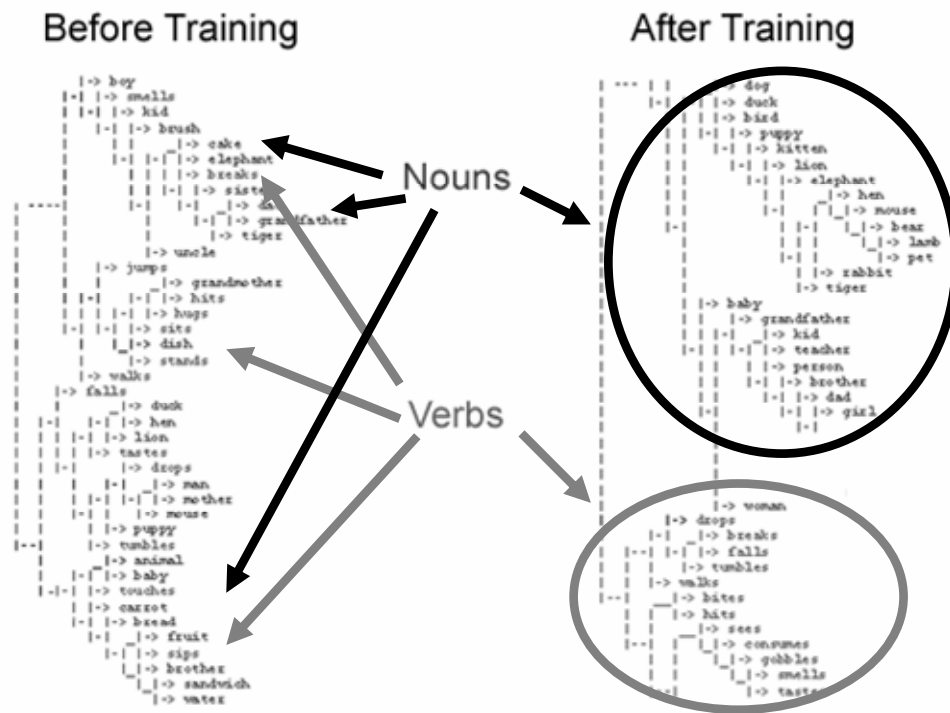


Figure 1-2. Hidden Unit representation of vocabulary items in a young network before extensive training, after 20,000 sweeps and then at the end of training at 140,000 sweeps

other, it is possible then to measure the “global coherence” (Keibel & Elman, 2004) of the category members.

Keibel and Elman (2004) compares two different algorithms to measure these activations, and find that average precision(Zavrel, 1996a) is appropriate in cases where the number of members between categories differ. Average precision values are then to be

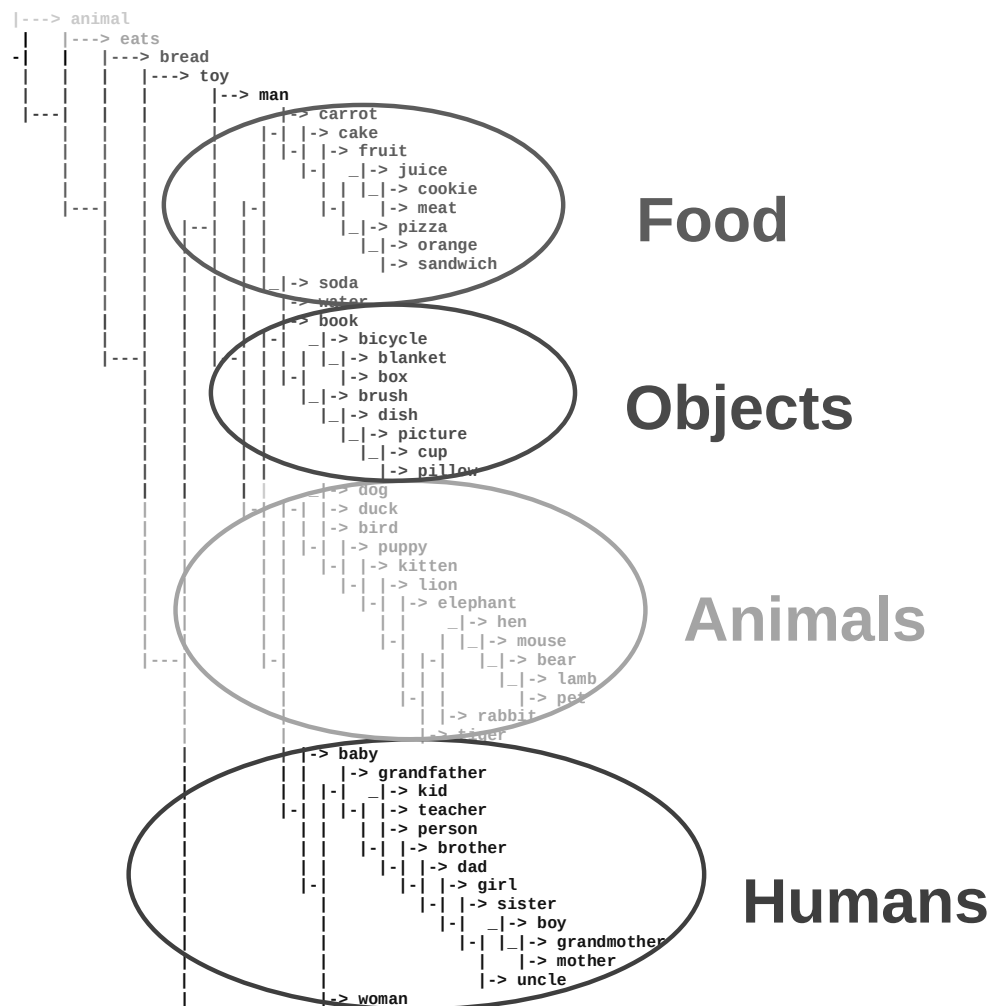


Figure 1-3. Close-up of hidden unit cluster of noun items in older network

used in this project, since one of the main questions of interest is how quantity and quality input may affect categorical formation. Average precision (AP) values range on a scale between 0 and 1, with higher values signifying that members of a category have more similar hidden unit activation values. Essentially, categories that are well-formed should have higher AP than those that do not.

1.1.5 Main questions: Global experience

The issues outlined above motivate a series of simulations that examine the influence of global linguistic input on new word learning and category development. Three simulations are reported that compare 1) the effect of input quantity, 2) the effect of token frequency, and 3) the effect of structural complexity.

1.1.5.1 Effect of Quantity

The amount of input has been shown to play a role in lexical acquisition, and the appearance of categorization abilities seems to be coincident with improvements in rates of lexical acquisition. If these two are related, then simulations should show that networks that receive larger amounts of input also will have higher category coherence. At the same time, networks with larger training inputs should also be able to learn new words more quickly. This would lend further evidence to the idea that improved linguistic learning is related to cognitive development.

1.1.5.2 Effect of Frequency

In order to fully examine how word learning might be influenced by category coherence, it is necessary to create conditions that are very similar, but one results in higher coherence values than the other. A recent proposal by Goldberg, Casenhiser and Sethuraman (2004) suggests one way this may be possible. In this corpus study, this group finds that there are five highly frequent verbs in CDS that correspond to five common constructional categories. This is followed an empirical work with both adults (Goldberg, Casenhiser, & Sethuraman, 2004) and children (Casenhiser & Goldberg, 2005) that observe that verb construction learning is improved when the construction has been learned with a

highly frequent exemplar. This idea suggests that networks also might form a particular category such as ANIMAL more readily if they are exposed to input in which one animal word, like cat, is more frequent than other animals. Under this hypothesis, networks that are exposed to all words in category with equal frequency should form less coherent categories. As a result, novel word learning in that category should occur more slowly, compared to categories that were trained with highly frequent exemplars.

To summarize, two predictions are made. First, networks will form more coherent categories when trained on input that contains a one highly frequent word and other low frequency items, rather than several words of equal frequency. Second, lower coherence values should lead to slower novel word learning, compared to networks with higher coherence values.

1.1.5.3 Effect of Syntax

As described earlier, there is conflicting evidence on the relationship between syntactic complexity in CDS and early lexical learning. While some scientists have observed that shorter, simpler utterances are helpful in acquiring early words, others have reported that more complex speech predicts higher vocabulary levels in infants. It is possible that this conflicting evidence arises from some of the disadvantages of CDS measurement in toddlers. One drawback is that data are often collected sporadically in lab visits, or, in cases where children are recorded many times and at home (Cameron-Faulkner, Lieven, & Tomasello, 2003), the sheer amount of the data precludes having a large number of participants from a variety of backgrounds. Additionally, it is not possible to systematically alter the language input a child may hear on a large scale. On the other

hand, it is possible to know in detail about the entirety of experience a neural network has with language, and to experimentally vary important properties of this input.

In this case, the most straightforward comparison of grammatical complexity in input is to present networks either with only simple, transitive and intransitive sentences or networks that contain this simple input plus more complex ditransitive and matrix sentence constructions. If this additional grammatical complexity hinders vocabulary growth, then networks trained with the latter type of input should learn new words more slowly than networks that have been exposed to only more simple constructions. Second, this slower word learning should then be tied to reduced categorical coherence.

1.2 Effects of local experience on word learning:

In the sections above, I have reviewed a number of issues relating to the influence of global linguistic experience on lexical and categorical development. A number of questions arose from this review that will be answered via computational studies in Chapter 2. The review made clear that linguistic experiences accumulated over a lifetime can influence subsequent lexical learning. However, when learning a word, the surrounding context in which it appears is also critically important. For example, it is possible to imagine contexts that would be more or less helpful when understanding an unknown word's meaning. A sentence like "There is a foop," would not be as informative as, "I wanted to start the car, but couldn't find a foop." In fact, when learning words in helpful contexts, it is possible to acquire the meaning of novel words very quickly, sometimes after a single instance. How do our brains accomplish this feat? What kind of

information do our brains acquire about these rapidly learned words? I present a set of three studies that aim to explore these questions.

Unlike the experiments discussed in the first section on global experience, I examine contextual word learning in adult participants in this section. This may seem unusual since the majority of word learning research explores learning in pre-literate children learning their first language, or adult bilingual populations. In general, these studies examine word learning in explicit object learning paradigms. However, the majority of an individual's vocabulary is neither acquired explicitly, nor during early childhood. Rather, the greater part of an individual's vocabulary is acquired after entering school and learning to read. For example, estimates of adult vocabulary levels vary widely, but generally fall between 40000–150000, whereas a pre-literate five or six year old will know only 2500-13000 words (Aitchinson, 1994; Beck & McKeown, 1991; Bloom, 2000; Pinker, 1994). It is clear that adults learn a tremendous number of words.

Moreover, adults and older children learn words differently from younger children. Pre-literate children often learn new words through explicit naming and reference, while school age children and adults acquire words almost entirely incidentally in various language contexts, especially reading (Jenkins, Stein, & Wysocki, 1984; Nagy, Anderson, & Herman, 1987; Nagy, Herman, & Anderson, 1985; Sternberg, 1987). This mode of learning can be remarkably fast. Under the right conditions, only a single exposure to a novel word may be sufficient for a learner to infer its probable meaning (Carey & Bartlett, 1978; Dollaghan, 1985; Heibeck & Markman, 1987). Nevertheless, relatively little is understood regarding the role of linguistic context in word learning, what information is rapidly learned about word

meanings, and even less has investigated the cognitive and neural underpinnings of this ability.

This section of the dissertation addresses a number of questions regarding contextual word learning which are addressed through several studies conducted with normal college participants. The main goals of this dissertation are to 1) delineate how one factor in sentence context – contextual constraint, influences lexical acquisition, 2) probe what kind of knowledge is rapidly learned from context about word meanings, and 3) better understand the neural correlates of this ability.

To accomplish these goals, a series of three event-related potential (ERP) studies have been conducted. The first goal – to understand how sentence context influences lexical acquisition, is addressed by all three studies. In each study, novel words are presented in sentences that either strongly or weakly constrain the possible meaning of the novel word. After a single opportunity to learn novel words from these sentence contexts, the knowledge of the acquired novel word meanings is measured. Experiments one and two are designed to address the second goal – to examine what kind of knowledge is rapidly acquired about word meanings. Experiment one probes novel word knowledge via a plausibility judgment task that assesses knowledge of appropriate word usage, whereas the probe task in experiment two is a semantic priming paradigm designed to address understanding of the relationship between newly learned words and other already understood words. The third goal, which is to explore the neural correlates of word learning, are addressed in all three studies. The first two studies examining how neural activity measured at the scalp is modulated as a result of meaning acquisition. Experiment three then documents the representation of novel word meanings in the right and left hemisphere.

Before providing a more detailed description of the studies, the following section motivates the experiments with a review on the state of knowledge relevant to these stated goals in two relevant areas: 1) the neural correlates of word learning in both adults and children, 2) the role of context in word learning. I identify how it may be possible to measure lexical acquisition from brain activity measured via electrical activity at the scalp. But first, as all of the studies in this dissertation are conducted using the event-related potential technique, I provide an introduction of the methodological background and assumptions of this approach.

1.2.1 Event-related potentials

Event-related brain potentials, simply put, are averaged electrical brain responses that are time-locked to an “event.” These electrical brain responses are measured via electrodes placed on the scalp, and are thought to reflect sustained post-synaptic potentials produced by large populations of pyramidal neurons that are spatially aligned with each other and are perpendicularly aligned with the cortical surface. Since the measured ERP activity can be generated by a nearly infinite combination of various neuronal populations of varying size, strength and distance from the electrode site(s), it is difficult to determine the precise cortical/brain regions that are involved in the production of ERPs. On the other hand, the ERP technique has high temporal resolution - up to 1ms or less under certain conditions. This lends ERPs particular utility in answering questions related to the time-course of cognitive and sensory brain activity. In addition, ERP methods have advantages over pure behavioral measures as they can measure differences in cognitive processing that may not always be reflected by many behavioral measures.

For instance, in a reaction time task it is impossible to distinguish if identical performance between two conditions is mediated by identical cognitive processes or by two different processes which happen to result in identical motor performance. In contrast, ERP measures can often distinguish between cognitive processing differences that are not reflected by reaction time. However, since the measured ERP activity at any time point also reflects numerous other non-event-related processes, many more trials are needed (ranging from a couple dozen to hundreds) to yield the final event-related potential waveform in order to factor out the effect of random/unrelated activity. A typical ERP waveform consists of a series of positive and negative voltage deflections, referred to as components. Early components are sensory responses known as exogenous components. Later, endogenous components - those that are elicited more than 200-300 seconds after stimulus onset, are thought to reflect internal factors in cognitive processes, as opposed to responses to external stimuli properties. An example of an ERP waveform at a single electrode in Figure 1-4 with a number of labeled components. In this figure, negative amplitudes are plotted up, and time after stimulus onset is plotted at 100 ms tick mark intervals on the X-axis. This figure shows an initial positive deflection, known as the P100, followed by a negative peak (the N100), followed by a larger positive deflection, the P200, and then a large negative wave, known as the N400. In general, components are named with "N" or "P" reflecting their amplitude, and a number reflecting their approximate peak time after stimulus onset.

A number of ERP components have been documented as involved in language comprehension and processing, including the N400, P600 and LAN. The N400 is a negative going wave that peaks approximately after the onset of any potentially meaningful

stimulus (Kutas & Hillyard, 1980). This includes both linguistic stimuli such as written and spoken words presented both individually and in sentences, as well as for non-linguistic stimuli such as pictures, environmental sounds, gestures and videos. Its distribution tends

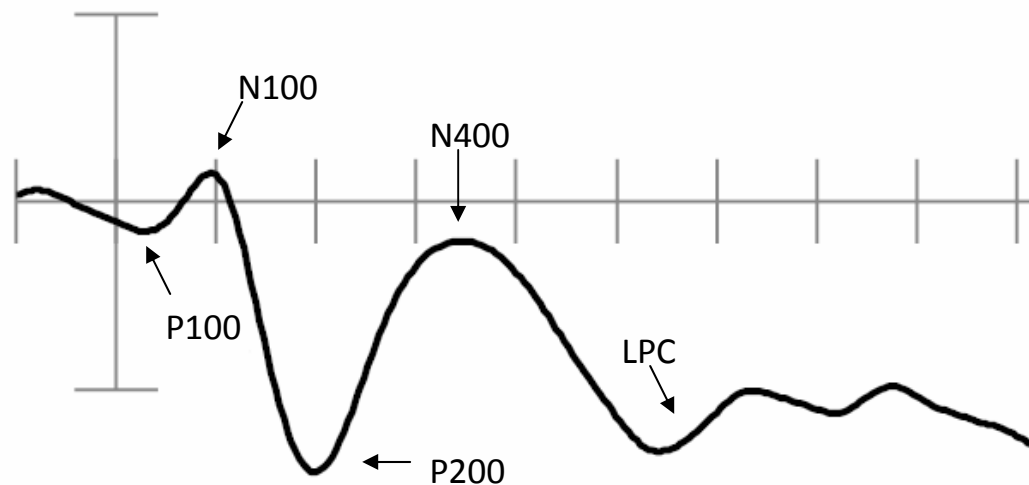


Figure 1-4. Example ERP waveform plotted at a single electrode with a number of common components labeled. Negative is plotted up, and time from stimulus onset is plotted across the X-axis in 100 ms intervals, starting with a 100ms pre-stimulus baseline.

to be maximal around centro-parietal electrode sites, amplitude has been found to decrease when a word is more expected or when features associated with its meaning are more easily integrated within its surrounding context (Federmeier & Kutas, 1999; Kutas & Hillyard, 1980; Van Petten & Kutas, 1990). For example, Kutas and Hillyard (1980) recorded brainwave responses to sentences completions that were either congruent or incongruent with the context of the preceding sentence. In a sentence like: "I drank my coffee with cream and milk" where the sentence ended was congruent given the sentence

context, the evoked N400 response was much lower than to sentences like “I drank my coffee with cream and dog” where the sentence completion was not congruent to the sentence context. It has also been found that one of the best predictors of N400 amplitude is cloze probability (Kutas & Hillyard, 1984). Cloze probability is measured by determining the probability that a particular word is given in a context on a sentence completion task. For words with small cloze probabilities, the N400 is large, and the N400 decreases accordingly as cloze probability increases. This suggests that the N400 is not only enhanced to semantically anomalous words, but also is related to a word’s degree of expectancy or ease of which its meaning may be integrated with the surrounding context. Additionally, the N400 for pronounceable nonwords is larger than that of real words, but is not present for nonwords that do not have orthographically legal spellings, or are unpronounceable (Bentin, 1987; Kutas & Hillyard, 1980). Therefore, N400 amplitude is associated with how meaningful a word is in a given context, ranging from very small when a word is very easily integrated or understood, to very large when the meaning of a word is unknown.

Syntactic violations in sentences can also elicit effects in two unique ERP components, the P600/SPS (syntactic positive shift) and LAN (Left Anterior Negativity; Coulson, King, & Kutas, 1998; Friederici, Hahne, & Mecklinger, 1996; Kluender & Kutas, 1993; Osterhout & Holcomb, 1992). The LAN has a latency similar to that of the N400 – but has a left anterior distribution, and the P600/SPS is a slow-going positive shift elicited approximately 500 ms as onset of the violation, and can last for 500-600ms. While the underlying mechanisms involved in each of these components is debated (see Coulson, King, & Kutas, 1998), they are elicited by a broad range of syntactic violations, including

those of grammatical agreement such as case, gender, and number, as well as violations of phrase structure and subcategorization preference (see Hagoort, Brown, & Osterhout, 1999 for a review) .

In sum, the ERP methodology has been established as a highly useful instrument in the cognitive scientist's toolkit that has helped to gain a better understanding of various cognitive mechanisms involved in language processing. In this dissertation, I will describe how one particular ERP component – the N400, can be used as an index of word meaning acquisition, in order to better understand how sentential context influences the kind of knowledge rapidly acquired about word meanings. In addition, I use this component and method to explore how the brain represents information about newly learned words. Below, I review the current state of knowledge concerning the neural mechanisms of word learning.

1.2.2 Neural correlates of word learning

Research on the neural correlates of word learning has recently become a topic of much interest, especially for investigators studying word learning in toddlers (i.e. (Friedrich & Friederici, 2004; Friedrich & Friederici, 2005a, 2006; Mills, Coffey-Corina, & Neville, 1997; Mills, Coffey-Corina, & Neville, 1993; Mills, Plunkett, Prat, & Schafer, 2005; Torkildsen et al., 2006; Torkildsen, Syversen, Simonsen, Moen, & Lindgren, 2007) and in adult second language learners (McLaughlin, Osterhout, & Kim, 2004; Osterhout, McLaughlin, Pitkanen, Frenck-Mestre, & Molinaro, 2006). Relevant studies in toddlers have explored brain responses to known and unknown words (Mills, Coffey-Corina, & Neville, 1997; Mills, Coffey-Corina, & Neville, 1993), or to words that are paired with congruent or incongruent picture contexts

(Friedrich & Friederici, 2004; 2005a, 2005b) or the role of experience and vocabulary knowledge upon lateralization of brain activity to known, unknown, and recently learned words (Mills, Plunkett, Prat, & Schafer, 2005). Through this research, developmental researchers have observed the emergence of adult-like patterns of brain activity during lexical processing. Thus far, ERP studies have suggested that children as young as 14 months of age present with electroencephalographic (EEG) activation patterns that are similar to those of adults when encountering known words in congruent and incongruent contexts (Friedrich & Friederici, 2005a). Additionally, lateralization of activity to words to the left hemisphere develops into an “adult-like” form relatively early, between the ages of 13 and 20 months (Mills, Coffey-Corina, & Neville, 1997), and that by 20 months of age, toddlers brains are able to distinguish between words that are related and unrelated in meaning (Torkildsen et. al. 2006; 2007). Since toddlers also show dramatic gains in word learning ability around this time, it is likely that the changes in this period mark a developmental shift in which children begin to use adult-like cognitive and neural mechanisms while processing unknown words.

At the same time, it has been observed that most adults do not possess the same degree of proficiency as child language learners. The mechanism responsible for this pattern is still debated. Some theorists have proposed the presence of a “critical period” for language learning whereby the processes involved in language learning as a child become unavailable to adult learners (Lenneberg, 1967). Support for this idea can be found in a Johnson and Newport (1989) study which found that individuals who began learning English as a second language after puberty scored lower on several tests of grammatical ability than those who learned before puberty. Additionally, this effect was unrelated to the number of

years of experience with English. On the other hand, more recent studies have observed a continuous decline in language proficiency with age that is not associated with the onset of puberty (Hakuta, Bialystok, & Wiley, 2003). These findings suggest that the ability to learn a second language declines gradually with age, rather than declining dramatically around puberty (Hakuta, Bialystok, & Wiley, 2003). This continuous effect has been replicated in a study that examined brain density of second language learners as a function of age of acquisition (Mechelli et al., 2004). This study finds that gray matter density of the inferior parietal cortex increases as an inverse function of age of acquisition of a second language. In addition to acquisition of a second language, more recently, age of acquisition of a first language has also been identified as an important factor in predicting level of language proficiency (Mayberry & Lock, 2003; Mayberry, Lock, & Kazmi, 2002). These studies examine language proficiency in individuals who have acquired sign language as a first language either natively, from birth, or have had a period of delay before acquiring a language in a school setting. They find that native speakers show improved results on tests of grammatical and lexical language ability over those who learned their first language only after entering school.

Proponents of a language critical period interpret the presence of adult-like patterns of brain activity during lexical processing in young children as evidence of more general semantic/cognitive language processing abilities that are retained into adulthood. On the other hand, recent adult studies have observed that only a few months of exposure to a new language in a college course is enough to result in brain responses to words in a new language that are indistinguishable from native speakers of the same language (McLaughlin, Osterhout, & Kim, 2004). In this research, participants who are enrolled in a college French course are asked to make lexical decisions on words and nonwords in French. They find that

students who have been enrolled in French for one year versus those who have only been enrolled for one month or three months show brain responses that distinguish between real and nonwords in French. The authors interpret this finding to mean that changes in neural activation in response to words can index changes in familiarity and knowledge of word meanings in a second language. Additionally, experienced second language learners and native speaker brain responses to real and nonwords were identical in this task, thereby demonstrating that the brain seems to process word meanings that are acquired in adulthood and childhood similarly.

It is interesting to note that Osterhout and colleagues (2006) have also observed that adult second language learners were not able to acquire native-like patterns of brain activity for novel syntactic features of a new language after a year of instruction. In this case, native French speakers and French students were tested on several syntactic features in French that were present either in both languages or just the second language. They found that features that were not present in English showed larger differences between native speakers and students than those that were present in English. Perhaps one reason for this is that lexical acquisition continues throughout the lifespan, whereas native syntax is completely mastered during childhood. Word learning is a lifelong task, which may also be why similar patterns of brain activity are observed in adults and toddlers during processing of unfamiliar words. It is thus not unreasonable to assume that studying the neural and cognitive bases word learning in college-aged students -- as is proposed herein -- is likely to inform us about the mechanisms involved in word learning in children as well.

EEG studies of word meaning acquisition have focused upon one particular brainwave component: the N400. As described in the previous section, the N400 amplitude

is associated with how meaningful a word is in a given context, ranging from very small when a word is very easily integrated or understood, to very large when the meaning of a word is unknown. This suggests that the N400 may vary with the degree to which the meaning of a newly encountered word is known. As described above, N400 amplitude in second language learners is reduced commensurate with experience with the second language (McLaughlin, Osterhout, & Kim, 2004). Mestres and Rodriguez-Fornells (2006) also reported that the N400 component amplitude changes as new words are repeated over several progressively more constraining sentences. It is therefore likely that improved knowledge of a word's meaning would be reflected in reduced N400 activity, but there is little data that speak directly to this prediction. Further, even less is understood as to what kind of knowledge the N400 represents about novel word meanings and how this information is represented in the brain. Therefore, a major goal of the experiments described in Chapters 3, 4, and 5 is to clarify the relationship between the N400 and word meaning acquisition.

Taken together, the above review demonstrates that the temporal dynamics of brain activity during lexical acquisition can reflect changes in knowledge of word meaning. However, spatial differences in brain activity may also uncover additional important information regarding the neural and cognitive mechanisms recruited during word learning. For reasons explained above, the EEG technique is not particularly well suited for fine-grained analyses of anatomical regions involved in cognitive processes. However, techniques like hemi-field presentation have been successfully employed to examine asymmetries in cognitive processing between the left and right hemispheres. Below, the following section provides a brief review of what is known about the neural areas recruited during word learning. This is followed by an overview of EEG studies that examine hemispheric

differences in lexical processing in order to identify how examining hemispheric asymmetries in word learning may shed light upon the cognitive and neural mechanisms involved.

1.2.3 Neural substrates of word learning

While there are some clues as to the temporal dynamics of word learning, only a few studies have examined the neural regions that may be recruited during word meaning acquisition (Breitenstein et al., 2005; Lee et al., 2003; Raboyeau et al., 2004). Generally, these studies have examined word learning that occurs over an extended training period, and measured changes that relate not just to meaning acquisition, but improvements in long term memory retrieval (Raboyeau et al., 2004) or increasing familiarity with a novel alphabet (Lee et al., 2003). However, Breitenstein and colleagues (2005) examined brain areas involved in relatively rapid association of novel words and objects over the course of ten word/object presentations. In this study, brain activation was recorded via fMRI as objects and auditory words were seen ten times in one of two conditions. In the first condition the objects and words were reliably paired together, and in the second they were not. During the task participants were asked to provide judgments as to whether the pairings were “correct” or “incorrect.” Modulations in brain activity associated with lexical learning were predominantly observed in the left hemisphere in the hippocampus, fusiform gyrus and inferior parietal lobe (IPL). However, since the focus of that study was on association of auditory words with picture objects over the course of ten presentations, it is unclear how these results would generalize to tasks that probe more immediate word learning that occurs with only a few exposures to a new word in linguistic context. Explicit object name learning in a first language has also been explored using magneto-encephalography (MEG)

(Cornelissen et al., 2004). Participants in this study were explicitly trained on the names and meanings of novel objects, and MEG activity to known, unknown, and trained objects was measured before, during and after the training period. Electrical dipole modeling revealed word learning related changes in activity that were associated with left IPL in 3 out of their 5 subjects. While the number of subjects in this study is small, and the MEG method is not as well suited as fMRI for studies of localization of brain activity, the involvement of the IPL area in word learning is consistent with findings in the word retrieval and word reading literature. The angular gyrus, which is a subregion of the IPL, has been implicated in word naming and reading abilities in fMRI and neuropsychology literatures as well. Additionally, this area is seated in a region between temporal, occipital, parietal lobes, and maintains interconnections between visual, auditory, and somatosensory association areas including: the middle basal temporal lobe, the superior colliculus (SC), the lateral geniculate nucleus (LGN), the frontal lobes, inferior temporal region, and other assimilation areas (Geschwind, 1965; Luria, 1973). Single neuron recording studies in this area have also found neurons in this area that are responsive to multi-modal inputs (i.e. neurons which fire in response to both audio and visual stimuli), which is consistent with the connective neuroanatomy of this region. Additionally, the functional specialization of this brain area is late to develop (Rivera, Reiss, Eckert, & Menon, 2005), potentially because its functionality is dependent upon connections from so many areas, and has been associated with later developing abilities like reading and arithmetic (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). Since it is responsive to multiple modalities, it is likely that this region may be important in integrating and retrieving information from several modal regions – which would be an important function in learning novel words from context. In addition to the angular gyrus there have

been a number of other regions associated with word retrieval, including broad areas in the left temporal lobe, such as the posterior medial temporal gyrus (Wernicke's area) and inferior temporal gyrus, as well as the inferior frontal gyrus (Broca's area) (review in (Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004). Neuropsychological and functional neuroanatomical literatures have identified Wernicke's area as an important area in language comprehension. This area maintains connections via the superior longitudinal fasciculus to Broca's area, which has been implicated in a number of language functions, including semantic integration and expression of speech. It is interesting to note that fiber tracts in the arcuate fasciculus also pass from Wernicke's area to structures within the IPL, including the angular gyrus. More broadly, the left inferior prefrontal cortex (LIPC) which encompasses Broca's area (which consists of Brodmann's areas (BA) 44 and 45), and extends through BA 47, has also been identified from neuroimaging (Gold & Buckner, 2002) and neuropsychological studies (Thompson-Schill et al., 1998) as important in semantic retrieval. Thompson-Schill and colleagues (1998) show that patients with damage in LIPC areas have more difficulty in generating verbs from nouns that have many associated verbs as compared to nouns that have only few associated verbs. This is taken to mean that the LIPC is involved in the selection of competing alternatives from semantic memory, rather than just semantic retrieval alone. From the review above, it seems likely that a network of regions that fall along the arcuate and superior longitudinal fasciculus white matter tract pathways are potentially involved in contextual word learning. Together, it is likely that these structures in the IPL bring together input from multiple modalities and integrate this knowledge into networks that are involved in language comprehension and production, and potentially, word learning.

However, while characterizing the neural time course and activation of lexical acquisition is an important task and just beginning to be studied, it is also important to understand the role of these areas in word learning and how they represent and process information about novel word meanings. The next section considers how a well-established EEG paradigm – the visual hemifield technique, can reveal important information about the representation of word meanings, and identifies how this technique can be applied to better understand the cognitive and neural correlates of word learning.

1.2.4 Hemispheric asymmetries in lexical representation

Left hemisphere (LH) dominance for language is one of the oldest and best known hemispheric asymmetries. Much of the initial evidence for this asymmetry comes from studies in patient populations who have received injury to a single side of the brain. These studies typically report profound deficits in language production and comprehension that result from left hemisphere damage, but few deficits due to right hemisphere injury. However, subtle communicative deficits have been observed after right hemisphere (RH) damage, such as in comprehension of indirect requests (Stemmer, Giroux, & Joannette, 1994; Weylman, Brownell, Roman, & Gardner, 1989), metaphors (Brownell, 1988), jokes (Brownell, Michel, Powelson, & Gardner, 1983) and familiar idioms (Van Lancker, 1987).

More recently, visual hemi-field paradigms have been used to explore hemispheric differences in processing in neurologically intact individuals. This method makes it possible to probe the relative role that a single hemisphere has in initial processing and comprehension of language, by presenting information to either the right or left visual fields (RVF or LVF) – which is then sent to the contralateral hemisphere .

Using this method, scientists have discovered that both cerebral hemispheres represent significant, albeit different, semantic and categorical information about word meanings (see Beeman & Chiarello, 1998 for a review). These studies have suggested that the RH is specifically involved in the processing of more distant, or weakly related semantic information. For example, ambiguous word meanings (RIVER-BANK vs MONEY - BANK) (Burgess & Simpson, 1988), weakly associated category members (DEER-PONY Chiarello, Burgess, Richards, & Pollock, 1990) and weakly related associates (Beeman, Friedman, Grafman, & Perez, 1994) are all more strongly primed in the RH. These hemispheric differences in semantic representation have been characterized as being “focused”, “fine” or “localized” in the LH and “coarse”, “diffuse”, or “distributed” in the RH (Beeman, Friedman, Grafman, & Perez, 1994; Chiarello, 1998; Grose-Fifer & Deacon, 2004). On the other hand, both hemispheres tend to show equivalent priming effects for more closely related or highly associated information (CAT-DOG; Chiarello, 1998 but cf.; Grose-Fifer & Deacon, 2004), as well as for repeated semantic information (Weems & Zaidel, 2005).

Despite the abundant evidence that both hemispheres participate in representation of known word meaning, nearly nothing is known about how the hemispheres come to represent novel word meanings. Developmental examinations of children with unilateral brain injury (Eisele & Aram, 1993; Thal et al., 1991), and ERP studies of early word knowledge in infants (Mills, Coffey-Corina, & Neville, 1997; Mills, Coffey-Corina, & Neville, 1993) have suggested that both hemispheres are important in early acquisition of word meaning. For example, Eisele and Aram (1993) measured single word naming and comprehension scores in children who had received LH or RH lesions at varying points during childhood and found lower scores in both groups on comprehension

and naming when compared to typically developing controls. However, when directly comparing the relative impairment in naming in comprehension due lesion side, children with RH lesions were relatively more impaired on comprehension measures than those with LH lesions. Similarly Thal et. al. (1991) measured early expressive and receptive vocabulary in a more uniform group of children who had received lesions at or around the time of birth via a parental report measure and found lower scores in children with both RH and LH lesions compared to typically developing controls. Once again, children with RH lesions were slightly more impaired than their LH lesion counterparts. In addition, early ERP measurements of known and unknown word processing in children between the ages of 13 to 20 months of age (Mills, Coffey-Corina, & Neville, 1997; Mills, Coffey-Corina, & Neville, 1993) has uncovered shift in laterality of brainwave response to known words from bilateral at 13 months to left-lateralized at 20 months, as well as increased left-lateralized responses in more children with higher vocabularies. This converging evidence from electrophysiological and lesion studies suggests that both hemispheres may mediate lexical learning in young children, and it is possible that the RH may have even greater involvement than the LH at especially early points in development.

Even though developmental work indirectly indicates involvement of both cerebral hemispheres in early lexical learning, there has been no work to directly assess this question in either children or adults. In Chapter 5, we examine the involvement of each hemisphere in the acquisition of word meaning. The experiment in Chapter 5 measures hemispheric representation of novel word meanings in adults (and not children) for a variety of reasons. First, the primary method that has been established to probe hemispheric representation of word meaning – hemifield presentation, is difficult to utilize

with children both behaviorally and electrophysiologically, due to the methodological necessity that participants remain still and fixated on a central point as stimuli are rapidly flashed to the left (LVF) or right visual field (RVF). With adults, it is also possible to utilize visual written stimuli for lateralized presentation, whereas studies with children more often use pictures or auditory stimuli that are difficult to accommodate in hemifield presentation techniques.

In this study, we ask how each hemisphere rapidly forms a representation of novel word meaning and how context can influence this representation by initially presenting adults with known and unknown words in written sentence contexts to both hemispheres simultaneously (central presentation), and then subsequently probing for acquired knowledge of word meaning via a lateralized semantic priming task. In addition, we explore how sentence context also influences this process by presenting novel words in sentence contexts that strongly or weakly constrain their meaning. In the priming task, known and unknown words appear centrally as primes for synonymous or unrelated target words that are laterally presented to the RVF/LH or LVF/RH. In this way, we can probe how each hemisphere represents words which have been newly acquired by adults via context.

1.2.5 Contextual Support and Semantic Knowledge

While a major goal of chapters 3, 4, and 5 are to examine the neural correlates of word learning, the central goal of these chapters are to explore the role of sentential context in word learning. Therefore, these chapters explore how immediate contextual information may affect novel word acquisition. In these studies, one component of contextual information is examined – that of contextual constraint. Novel words are presented in

sentences that vary in the degree to which a single meaning can be predicted from context. For example, a weakly constraining context such as “I walked across the room to Mike’s messy desk to return his *marf*” does not lead the reader to understand any specific meaning of “*marf*”. On the other hand, a sentence like “He put the pieces of the broken plate back together with *marf*” does suggest a single meaning for *marf* (glue). Despite the fact that the weak constraint context does not lead the reader to understand a single specified meaning of *marf*, it is still possible to glean some information about its potential meaning. What kind of information is understood about novel words from each of these contexts? It is possible that different kinds of semantic representations are formed due to the strength of constraint in which a new word appears. In Chapters 2 and 3, I describe two studies that examine how contextual constraint influences the acquisition of two important aspects of novel word meaning. The studies in Chapter 3 explore how constraint influences the understanding of subsequent novel word usage via a plausibility judgment task, and Chapters 4 and 5 examine the integration of novel words into semantic memory via a priming task. In Chapters 4 and 5, new words that initially appear in strongly and weakly constrained contexts are primed by words that are a synonym, a within category exemplar, or a between category exemplar of the intended meaning for the novel word. Previous studies have found that N400 amplitude to a target word varies with its relationship to a prime word (Bentin, Kutas, & Hillyard, 1993). Generally, less related primes that presumably share fewer semantic features with a target word elicit larger evoked N400 amplitude than primes that are highly related. However, in the case of learning a new word, it is unknown what kind of meaning is inferred in context, and how this might differ depending on the semantic constraint of the context in which it initially appears. Two possibilities are proposed: 1) Irrespective of contextual constraint, a

specific, basic level exemplar meaning is chosen for a new word; or alternatively, 2) Weakly constraining contexts result in a more general meaning of a word being chosen. There is evidence in support of both possibilities. In support of the first option, the developmental literature suggests that children tend to initially assign a basic level meaning to words, or learn basic level words first, such as “cat” or “ball” rather than more general, superordinate concepts like “animal” or “toy” (Brown, 1958; Markman, 1989). Additionally, children have been observed to learn new words very quickly – an ability coined as “fast mapping” of concepts (Carey & Bartlett, 1978; Dollaghan, 1985; Heibeck & Markman, 1987). This type of fast one-shot learning is consistent with the idea that a specific meaning is initially chosen for a word when given enough information about it initially. Furthermore, Federmeier and Kutas (1999) conducted an ERP study that presented participants with strong and weak constraint sentences with three types of completions: expected, unexpected within-category exemplars, and unexpected between-category exemplars. This study finds that N400 amplitude is larger to unexpected within-category sentence endings in weakly than strongly constrained contexts, suggesting that there is increased semantic facilitation of unrelated items in strongly constraining contexts (Federmeier & Kutas, 1999). This suggests that a weak constraint context would result in a narrower scope of semantic features being activated – and in the case of word learning, might suggest that a narrow scope of features would be assigned to a novel word that appears in a weakly constrained context, and vice versa. However, there is also support for the second option, in which meanings for words are acquired more gradually, and weak constraint results in a more general understanding of a word. For instance, behavioral semantic priming studies find that scope of semantic facilitation is increased in weakly constraining sentences (Schwanenflugel & Shoben, 1985).

In this research, participants completed a lexical decision task on sentence final words, where strong and weak constraint sentence completions were either an expected or unexpected word that was related to the expected word, similar in structure to Federmeier & Kutas (1999). Reaction times to expected and unexpected words did not differ in weak constraint contexts, but revealed facilitation for expected words in high constraint contexts. These results suggest that low constraint sentences might result in a broader scope of facilitation for upcoming items, while strong constraint narrows the scope of expected items, such that even similar items that are not highly expected show increased decision times in the LDT paradigm. Additionally, some educational researchers have suggested that words are gradually learned in degrees, where a more general scope of semantic features of a word is initially acquired (Schwanenflugel, Stahl, & McFalls, 1997). For example, Durso and Shore (1991) classified word knowledge into three categories: unknown, partial knowledge, complete knowledge, based upon whether participants could provide a definition for a word (complete), were familiar with a word (partial) or had never encountered the word previously (unknown). In a series of studies, they find that adult's ability to identify appropriate usage of words in these three categories varied in a continuous fashion. In this study, they found that participants were able to identify appropriate usages of unknown word at above chance level when a sentence using the word was contrasted with another. However, participants were unable to select the proper usage of unknown words from isolated sentences, whereas this was possible with partially known words. Additionally, adults were better able to fully learn the definitions of partially known words over unknown words (Durso & Shore, 1991). Schwanenflugel, Stahl & Mcfalls (1997) also follow up on this research in children and find that the amount of vocabulary growth in unknown and partially

known words is equal. This result lends support to the idea that word meaning is acquired gradually over multiple exposures, instead of a specific level of meaning being chosen at first, and then later the word knowledge is tweaked.

In conclusion, studying the neural correlates of lexical acquisition appears to be a fruitful area that will shed light upon neural and cognitive processes involved in language acquisition and extend our knowledge of the relationship between lifetime experience, and immediate context on word learning. In the following sections, I present a series of studies with the goal of advancing the field's understanding of these topics.

CHAPTER 2

LANGUAGE INPUT AND SEMANTIC CATEGORIES: A RELATION BETWEEN COGNITION AND EARLY WORD LEARNING

2.1 Abstract

Variations in the amount and nature of early language to which children are exposed has been linked to their subsequent language ability (e.g., Hart & Risley, 1995; Huttenlocher, Haight, Bryk, & Seltzer, 1991). In three computational simulations, we explore how differences in linguistic experience can explain differences in word learning ability due to changes in the development of semantic category structure. More specifically, we manipulate the amount of language input, sentential complexity, and the frequency distribution of words within categories. In each of our simulations, improvements in category structure, even when the nature of the input remains the same over time, are tightly correlated with subsequent improvements in word learning ability. These simulations suggest that variation in early language environments may result in differences in lexical proficiency by altering underlying cognitive abilities like categorization

2.2 Introduction

The ability to group objects into categories based on some similarity of function, form or meaning is arguably one of our most important cognitive behaviors. While it may be a common-sense notion that we form categories based on our own direct experience of them, we also develop categories for things we may have never directly experienced like Roman emperors, subatomic particles, and vacuum-tube computers. A reasonable assumption is that language plays a key role in making this possible. But a similar issue arises in the case of language: We ultimately are able to learn words for things outside our direct experience. This poses a chicken-and-egg problem. Intriguingly, during early stages of language learning there is little evidence of category knowledge, and the rate of vocabulary acquisition is slow. At the point when children undergo a ‘vocabulary spurt’—in which the pace of word learning increases rapidly—they also begin to display the ability to sort sets of objects into multiple categories (Gopnik & Meltzoff, 1987, 1992). This suggests that these two phenomena, i.e., the ability to learn new words and knowledge of categories, may be related in a synergistic fashion. In this paper, we use computational simulations to explore how language input influences development of category knowledge, and how category knowledge in turn influences subsequent lexical acquisition.

We begin with a brief review of the claims that have appeared in the literature regarding the nature of the relationship between category knowledge and lexical development. We then turn to a review of what is known about the effects of language input on vocabulary acquisition. Finally, we focus on a specific hypothesis—that even in the absence of experiential information, vocabulary acquisition can shape category knowledge,

which once in place, then facilitates the rate of learning new words—and study the conditions under which these two phenomena interact.

2.2.1 Relationships between lexical Development and category knowledge

How might the ability to categorize objects and the ability to learn new words be related? Several logical possibilities exist, ranging along a spectrum of being tightly interrelated to not related at all. The former position is suggested by Fodor (1983), who argues that linguistic and cognitive abilities are completely modular or domain specific and develop independently of each other. According to this account, semantic structure is not acquired due to language input. Rather, conceptual structure is innate, and words are learned that correspond to this innate 'mentalese.' Chomsky (1980) and Pinker (1991) have advanced similar—though not identical—positions.

But because there appears to be a confluence of rapid gains in cognitive and linguistic functioning around the middle of the second year, others have proposed that linguistic and other cognitive abilities, such as categorization, develop in a more tightly coupled manner. Within this camp, theories tend to differ on degree of this relationship, ranging from the strongly Whorfian (Whorf, 1956) hypothesis that thought is impossible without the use of language, to more interactive versions where language and cognition are completely integrated and rely on each other during development. We take this to be the position of, for example, Gopnik & Meltzoff (1993).

The possibility that word-learning might specifically drive conceptual organization has been proposed by Bowerman and colleagues (Bowerman, 1996; Choi & Bowerman, 1991). Bowerman cites evidence from cross-linguistic studies that children will readily learn the spatial categorization scheme present in their language. For example, Korean

focuses on the fit between objects, while English tends to emphasize the kind of containment or method of support. However, Choi & Bowerman (1991) find that children learning each language have no problem describing these spatial relations in terms appropriate to their language. So, Korean children will describe a peg in a hole, as tight or loose fitting, while English-speaking children will say whether the peg is 'in' or 'on' the hole.

Mandler (1996) has countered this claim by proposing that younger children might be able to carve-up space in ways that are not solely linguistically determined. She suggests that there are a number of different ways that the world can be categorized – all of which are initially available, but that language constrains these possibilities into a regularized convention that differs from language to language. Consistent with this claim is the finding that 0;9, 0;11 and 1;2 month olds all are able to distinguish between tight/loose and in/out distinctions, regardless of linguistic environment (Choi, McDonough, Bowerman, & Mandler, 1999). The above debate delineates two possibilities regarding the language-categorization interaction. The former proposes that language drives the kind of categorical structure that is formed. The latter maintains that non-linguistic categories are already formed from the infant's physical knowledge of the world, and that language 'slots' into these prior categories. By extension, one might assume that category knowledge facilitates vocabulary acquisition.

Lastly, Gopnik & Meltzoff (1993; , 1997) have argued that cognitive and linguistic development is related to each other in a more tightly interactive way, and that both can influence each other equally. For instance, they propose that exposure to a particular word will lead to a drive to learn the underlying concept, but that this new knowledge will then

interact with knowledge already in place, which might then lead to new word learning. Evidence for this idea is found from cross-linguistic studies in which the appearance of categorization abilities arise later in Korean speaking children than in English speaking children (Gopnik, Choi, & Baumberger, 1996). In Korean, verbs are much more prevalent than nouns, but the opposite is true in English. Since English speaking children are able to sort objects into categories earlier, this suggests that the language's emphasis on nouns drives the organization of (at least nominal) concepts into categories earlier than in Korean, and that this improved categorization ability is what facilitates subsequent word learning proficiency.

Logically, whatever the relationship between vocabulary acquisition and categorization abilities, the role of the actual input presented to a child must itself play a critical role in the acquisition process. We turn now to a review of what is known about the effect that input has on vocabulary acquisition.

2.2.2 The Role of Input

A growing body of research has found that early language input is key to predicting levels of lexical proficiency. For example, a number of studies (Huttenlocher et al, 1991; Hart & Risley, 1995) have found that children who have had more language input from their parents also know more words.

Additionally, some of the earliest observations about child directed speech (CDS) (e.g. Newport, 1997; Snow & Ferguson, 1977) have found that the language that children hear is simpler in both syntactic structure and the kinds of words used. For instance, one ambitious study of CDS to 12 children between the ages of 2;0 and 3;0 years old (Cameron-

Faulkner, Lieven, & Tomasello, 2003) found that 20% of all the utterances recorded during this period were sentence fragments, which usually were responses to a question, and that 32% of the utterances were questions. Complex sentences only accounted for 6% of all utterances heard by children at this age. These results were also compared with a similar, but smaller study by Wells (1981), who reported strikingly similar results.

These findings imply that simplifications in the structural complexity of CDS could aid word learning by reducing the processing demands placed on the child when a new word is encountered. However, in order to support this idea it is necessary to compare the kinds of input heard with the words actually learned by the child. To this end, Brent & Siskind(2001) examined the role of single word input in infants between 0;9 and 1;3 months on the words that the child knows at 1;6 months – i.e., at the beginning of the vocabulary spurt. They found that the number of times a child hears a word in isolation is a more reliable predictor of whether the word is known at 1;6 months than the total number of times the word is heard. By examining the role of isolated words, the investigators were able to study how the words heard in the simplest kind of grammatical construction affect the learning of that word. When taken into account with the other CDS findings described above, this work indicates that word learning as whole might be easier when the sentence structures remain simple.

On the other hand, there are also recent findings that more structural complexity in CDS improves word learning in 2;0 year olds (Hoff & Naigles, 2002). However, here syntactic complexity was not measured through detailed constructional analysis, but rather through the mean length of utterance (MLU) in maternal speech. Hoff & Naigles (2002)

concluded that it is more syntactic complexity, not less, that is important in aiding word learning.

In addition to differences in the structure of the input, there is also evidence that the distribution of words in input may differ amongst children (Bates, 1988; Broen, 1972). Weizman & Snow (2001) report that the usage of low frequency words varies between families, and that five year old children who encounter a higher proportion of this 'sophisticated words' from their environment also tend to have larger vocabularies. More recently, Pan, Rowe, Singer & Snow (2005) also find that variation in type and token frequency in maternal speech to children in the first three years of life also affects vocabulary growth. Moreover, they find that having a increased variation in maternal types was more significant than just overall amount of speech input alone.

We are thus left with a number of unresolved questions. That the input matters, and that there is some relationship between language development and category knowledge, seems almost trivially self-evident. But the specific effect of the input that a child hears upon vocabulary acquisition, and what the precise relationship between the word learning and category knowledge, remains unclear. Although the answers to these questions will ultimately come from empirical investigations, it would seem that there is a useful role to be played by using computer simulations to explore the process of lexical acquisition. In this way, it is possible to develop more specific hypotheses about the effect of different types of input, and the relationship between lexical acquisition and category knowledge.

Computational studies have been used in the past to model lexical learning (i.e. Li, Farkas, & MacWhinney, 2004; Plunkett, Sinha, Moller, & Stransby, 1992). These models

have been able to replicate phenomena that have been observed in lexical development of children. Of relevance to this paper, Li, Farkas & MacWhinney (2004) find that organization of lexical categories (nouns, verbs, adjectives and closed class words) in self organizing networks improve as the networks learn more words. This suggests that modeling can reveal important links between category and lexical development.

There are a number of ways in which computational simulations complement behavioral research. While it is difficult to devise a task that can directly measure internal representations in young children, and it would be unethical to alter language input to such an extent that it might disrupt normal processes of language acquisition, computational simulations can overcome these difficulties. This makes it possible to measure category structure both by artificially manipulating how well category members fit together, and more importantly, to evaluate the actual representations that form. That is, rather than inferring category knowledge from task behavior (as is done with children), the network's category knowledge can be assessed directly through analysis of its internal representations. We use artificially generated language input in order to have precise control over the properties of input, as it is difficult to find these kinds of neat and tidy kinds of variation in corpus data of CDS.

Another benefit of the computational methods employed in this paper is that it is possible to isolate the role that linguistic input alone may play in the development of conceptual structure, apart from other non-linguistic kinds of information that are undoubtedly used in lexical acquisition. In this way, our networks learn about words by their co-occurrence with other words. Previous computational simulations have demonstrated that such information can provide important information about category

structure (Elman, 1998; Elman, 1990). In the following simulations, we probe how categories that are learned in this way may be related to rate of word learning.

As a start, we propose that associations made from language input may alter underlying category structure, and that it is this change in category structure that can be related to proficiency in word learning. Under this hypothesis, category development is an important factor in word learning, so it can be expected that factors in language input that affect lexical learning outcomes, should also be reflected in the development of category structure similarly. This leads to two key predictions:

- 1) Variation in language input that affects lexical acquisition also affects development of category coherence in a similar manner; and
- 2) It is the development of categories themselves, and not solely language input, that facilitates lexical acquisition.

In the remainder of this paper, we examine this hypothesis and its predictions in three computational simulations with connectionist networks. In the first study we examine the role of the amount of input in developing category structure and subsequent acquisition of new words. The second investigates the role of syntactic complexity on this relation. The third experiment examines the role that distribution of word frequency in affecting this relationship.

2.3 The Modeling Task

The purpose of the following simulations is to explore the ways in which language experience might affect cognitive development, and how such development might in turn impact word learning ability. To this end, the network's category knowledge is assessed

directly through analysis of its internal representations. Before proceeding to describe the simulations, we provide a brief explanation of the neural network model that is used and explain how this type of model allows us to measure development of category structure.

2.3.1.1 Simple recurrent models

In this paper, we use the Simple Recurrent Network architecture (SRN, Elman, 1990). Simulations were run using the TLEARN software (Plunkett & Elman, 1997). This type of network is especially useful for processing elements that are sequential in nature, such as words in a sentence. Figure 2-1 illustrates the architecture of an example network. In our simulations, the network receives individual words as input, one at a time. The network's task is to predict the next word in the sentence as its output. The output that is produced is a function of both the current word input to the network and the prior internal state of the network stored in the context layer (Figure 2-1c). Importantly, this prior internal state is not a literal tape recording of preceding words, but is rather an abstract representation—that must be learned—of that sequence. The hidden layer that reflects these internal states (Figure 2-1b) is the part of the network that will be analyzed to provide evidence the network's knowledge of category structure, as described below.

2.3.1.2 Measurement of category structure

The use of language-like input allows for examination of the kinds of representations that might form when words are related by their occurrence in similar contexts to other words in the category. This involves active probing of the network to measure the coherence of representations that the networks have developed from this kind of input. More specifically, measurement of category formation is accomplished through calculation of the

network's internal (hidden unit) representations of words in a particular category (see Figure 2-1 for an example network). Here, words that the network has learned to be similar (more precisely, in the sense that they share similar linguistic properties), or to belong to a similar category, will share hidden unit activations that are more similar than those that are not within a learned category. For example, it is possible to cluster these representations graphically (Figure 2-2 and Figure 2-3) to visually reveal the similarity of hidden unit values for each word. Figure 2-2 outlines how hidden unit representations change over training in this study. Before the network has adequately learned about category structure, words are clustered without any noticeable relation to each other. Yet, at the end of training it is clear that the network has learned not only the differences between nouns and verbs, but also subcategories between them (Figure 2-3). By comparing how close all members of a category are to each other, it is possible then to measure the 'global coherence'. Here, we follow (Keibel & Elman, 2004) use of Average Precision (Zavrel, 1996b) for calculating global coherence. In this technique, the vector distance between each pair of words is calculated, and then for each word, other words are ranked on this distance measure. A coherence score is then calculated for each word based on how close other members of its category are ranked. The coherence score is then averaged across all members of a category to determine the overall coherence of the category. As Keibel and Elman (2004) point out, Average Precision (AP) is appropriate in situations where the number of members in different categories is not the same, which is the case in this study. AP values range on a scale between 0 and 1, with higher values signifying that members of a category have more similar hidden unit activation values.

Essentially, categories that are well- formed should have higher AP than those that do not (see Appendix 1 for a detailed discussion of average precision).

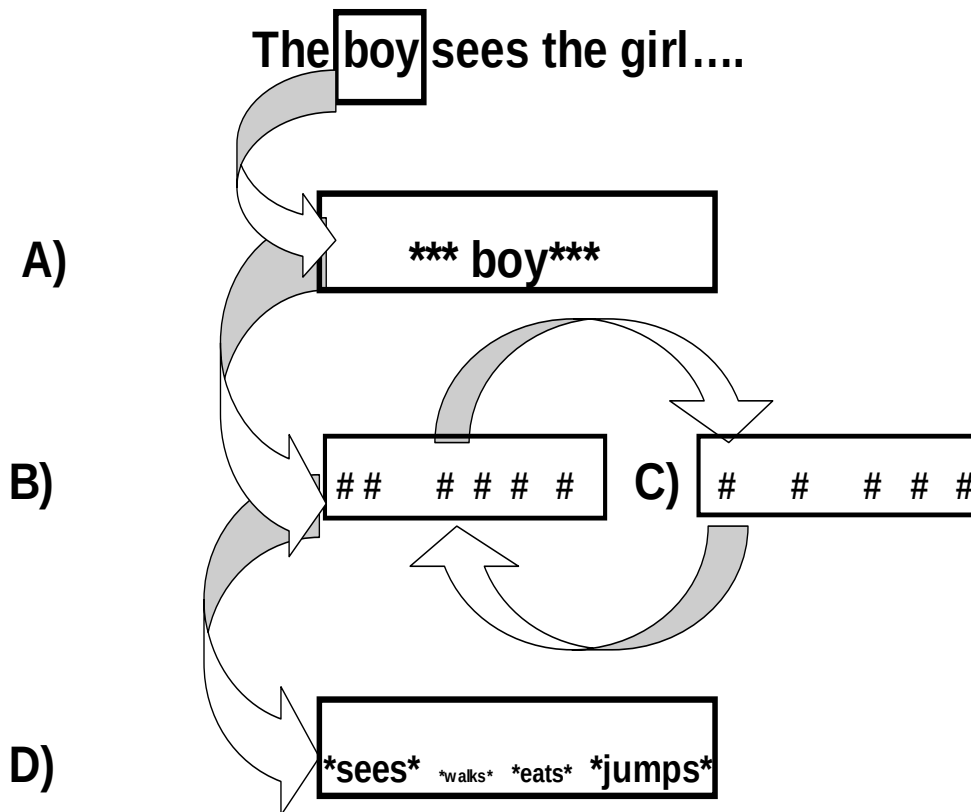


Figure 2-1. Above, words from the sentence "the boy sees the girl" are presented one word at a time to the network to the input layer (A). Next, the word is fed into the hidden layer (B), which also receives information about the immediately previous network states from the recurrent layer, (C). As training proceeds, the hidden layer will develop numerical internal representations that can be used to generalize to similar inputs. In the output layer (D) the network predicts the next possible words after 'boy'. Generally, the SRNs will learn to predict a range of words that are possible that correspond to their frequency with which they are associated. Here, the network is predicting both the actual next word "sees" but also other possible words like "jumps" and "eats".

The use of computational simulations also allows for identical networks to be exposed to different training environments. In this case, we alter the exposure to both the quantity of input, but also qualities of its structure and frequency of words within a category, as outlined in the next three sections.

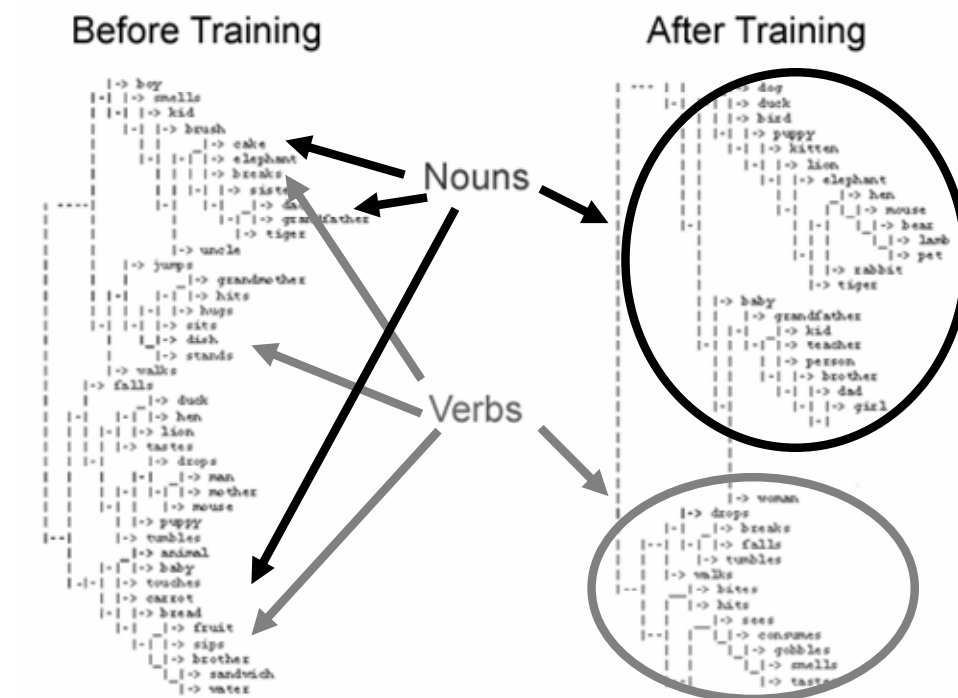


Figure 2-2. Hidden Unit representation of vocabulary items in a young network before extensive training, after 20,000 sweeps and then at the end of training at 140,000 sweeps

2.3.2 Effect of Quantity

The amount of input has been shown to play a role in lexical acquisition, and the appearance of categorization abilities seems to be coincident with improvements in lexical acquisition. If these two are related, then simulations should show that networks that

receive larger amounts of input also will have higher category coherence. At the same time, these networks should also be able to learn new words more quickly. This would lend support to the idea that improved linguistic learning is related to category development. This hypothesis can be further explored by manipulations of the input that more directly but subtly affect category development.

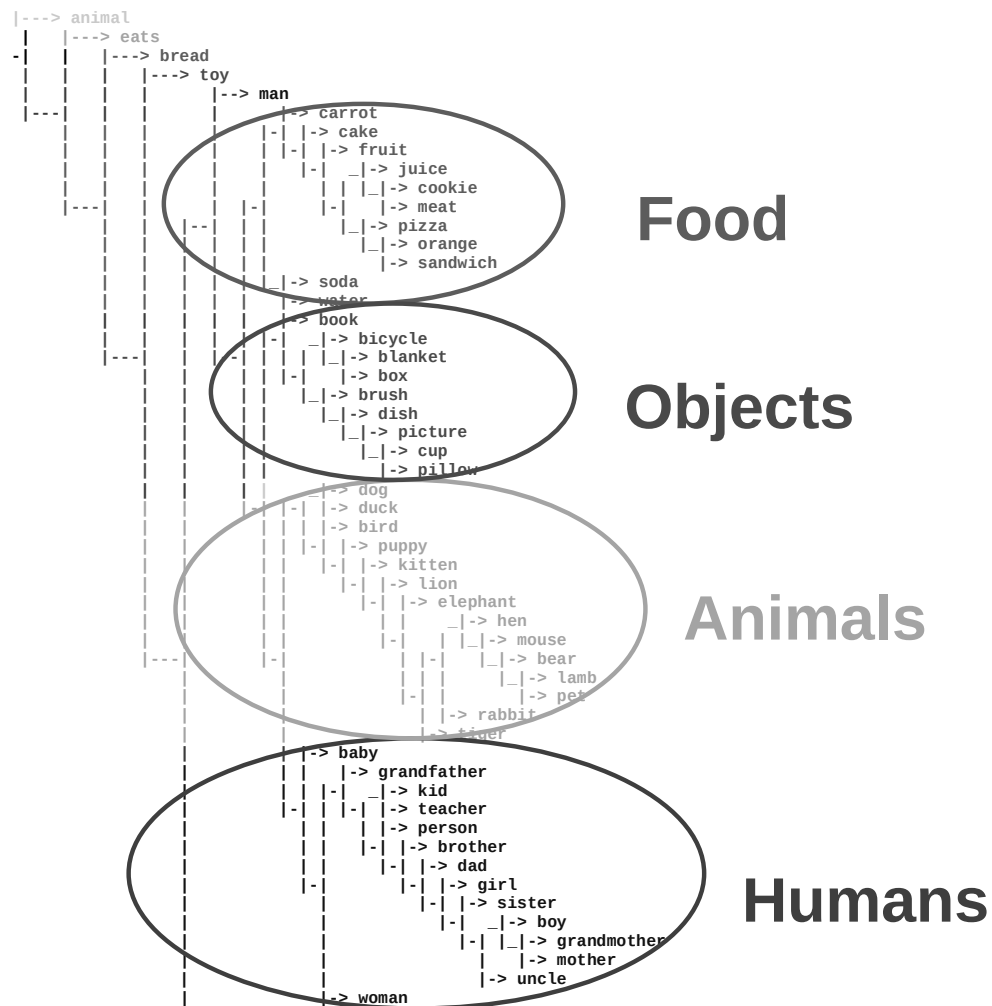


Figure 2-3. Close-up of hidden unit cluster of noun items in older network

2.3.3 Effect of Frequency

In order to fully examine how word learning might be influenced by category coherence, it is necessary to compare input conditions that are very similar, but in which one leads to higher coherence values than does the other. A recent proposal by Goldberg, Casenhiser and Sethuraman (2004) suggests one way this may be possible.

In that work, a corpus study revealed that there are five highly frequent verbs in CDS that correspond to five common constructional categories. Furthermore, an accompanying experimental study found that when new verb constructions are taught with a highly frequent exemplar, novel verbs with the same meaning construction are learned more easily for both adults and children (Casenhiser & Goldberg, 2005). This suggests that networks also might form a particular category such as ANIMAL more readily if they are exposed to input in which one animal word, such as cat, is more frequent than other animal words. Conversely, networks that are exposed to all words in a category with equal frequency should form less coherent categories (at least initially), and thus, when given a task to learn a new category member word, should learn less quickly than networks that have been induced to form a more coherent category through skewed word frequency exposure.

To summarize, two predictions are made. First, networks that are exposed to input where there is one very frequent word per category should form more coherent categories than those that have no frequency differences between words in the category. Second, networks that form categories with lower coherence, as measured by AP values, should learn new words in the low coherence category more slowly than networks with higher coherence values for a category.

2.3.4 Effect of Syntax

Finally, as discussed earlier, there is some uncertainty in the literature about whether more or less syntactic complexity in CDS is better for early language learning. One drawback in studies of speech to toddlers is that data are often collected sporadically in lab visits, or, in cases where children are recorded many times and at home (Cameron-Faulkner, Lieven, & Tomasello, 2003), the sheer amount of the data precludes a having large number of participants from a variety of backgrounds. Additionally, the success of these studies hinges upon that amount of variation that can actually be observed through recording sessions since it is not possible to systematically change the kinds of language input a child may hear on a large scale. On the other hand, it is possible to know in detail about the entirety of experience a neural network has with language, and to experimentally vary important properties of this input.

Of course, what counts as grammatical complexity is itself a complex question and one can imagine many ways in which utterances might be judged to be more or less complex. In this case, the most straightforward manipulation that lends itself to examination of how grammatical complexity of input may affect word learning is to present networks with input that contain only simple, transitive and intransitive sentences or networks that contain this simple input plus more complex ditransitive and matrix sentence constructions. If this additional grammatical complexity (as defined in this very specific manner) does indeed hinder vocabulary growth, then networks trained with the latter type of input should learn new words more slowly than networks that have been exposed to only more simple constructions. Second, this slower word learning should then be associated with lower AP values.

2.4 Experiment 1: Effect of Quantity

2.4.1 Method

2.4.1.1 Input

The input for all simulations that are described in this paper was constructed using a language generator program (SLG; Rohde, 1999). In the Experiments 1 and 2, a vocabulary of 85 words (52 nouns and 33 verbs) was used and sentences were formed corresponding to two simple syntactic constructions containing only nouns and verbs: NV and NVN (intransitive and transitive, respectively).

Nouns were assigned to the following semantic categories: ANIMALS (15), HUMANS (15), FOOD (12), and OBJECTS (10). Nouns in these categories are commonly observed in the early vocabularies of toddlers (i.e., Nelson, 1973), and are included on checklists for vocabulary checklists at this age (Fenson, Dale, Reznick, Bates, & Thal, 1994). Verbs belonged to the following categories: CHANGE OF STATE (5), COMMUNICATION (6), MOTION (6), EATING (5), PERCEPTION (6), and ACTION (5). These verb categories were chosen because they are typically included on checklists for vocabulary (Fenson, Dale, Reznick, Bates, & Thal, 1994).

In the artificial grammar that was used, nouns and verbs were required to agree both semantically and syntactically, meaning that a sentence had to be grammatically correct and ‘make sense’. Each word was coded in a localist fashion as an 85 element binary-valued vector with each bit representing a distinct word. Appendix B contains

appropriate categorical semantic relations between sentences. Appendix C contains examples of sentences used in the study.

The amount of input was manipulated by altering the numbers of sentences to which the network is exposed. Training involved input ranging between 20 and 1000 sentences, depending on condition; there were five conditions with corpus sizes of 20, 50, 100, 200, 500 and 1000 sentences respectively. Table 2-1 contains information about the number and kinds of types and tokens for each corpus size.

Although the relationships between categories was fairly simple, the largest input condition (1000 sentences) still presented the network with only a small subset of all possible sentences possible in this grammar. It is estimated that in order to see every possible sentential combination the network would have to see nearly half a million sentences. By training with only a subset of all possible data, this allowed for the network to be exposed to both a range of types and tokens, but still not see every possible combination, such that category membership was not plainly 'given away'. Instead the network was forced to generalize from incomplete input in order to figure out which words belong to which categories.

2.4.1.2 Training

For each of the input conditions listed above, 10 simple recurrent networks (SRNs, see Figure 2-1 for an example) with 50 hidden units were trained on a next-word prediction task. In this task, the network is presented with successive words in a sentence, one at a time, and is trained to predict the next word. Because the task is non-deterministic, the network's optimal strategy should be to learn the implicit

Table 2-1. Number of word types in each condition

No. of sentences	Condition			
	Even	Uneven	Simple	Complex
1000	85	85	85	95
Verbs	33	33	33	43
Animals	15	15	15	15
Humans	15	15	15	15
Food	12	12	12	12
Objects	10	10	10	10
500	85	--	--	--
Verbs	33	--	--	--
Animals	15	--	--	--
Humans	15	--	--	--
Food	12	--	--	--
Objects	10	--	--	--
200	79	--	--	--
	32	--	--	--
	14	--	--	--
	15	--	--	--
	10	--	--	--
	8	--	--	--
100	62	--	--	--
	25	--	--	--
	11	--	--	--
	12	--	--	--
	7	--	--	--
	7	--	--	--
50	44	--	--	--
	19	--	--	--
	11	--	--	--
	6	--	--	--
	4	--	--	--
	4	--	--	--
20	33	--	--	--
	15	--	--	--
	5	--	--	--
	6	--	--	--
	4	--	--	--
	3	--	--	--

classes of words that are appropriate in each context. Note that these categories are defined solely by privilege of occurrence (i.e., the input vectors themselves contain no information regarding category membership).

The learning rate was 0.01, no momentum was used, and training was carried out to 140,000 sweeps (one sweep corresponds to presentation of one word; pilot studies had determined that was sufficient training to result in asymptotic performance on the task). Thus, each network saw an equal number of words, but was exposed to each corpus a differing number of times, depending on the size of the corpus. This training scheme was meant to simulate a number of children who are the same age, but have heard different amounts of language input. Therefore, all networks had the benefit of being trained for an equal duration. The only difference was the range of sentences seen.

2.4.1.3 Analysis

At intervals of 20,000 sweeps, each network was probed to assess its ability to learn new words and also the category structure of its internal representations (i.e., hidden unit activations in response to seeing each word) as measured by AP scores.

New word learning was measured by exposing the network to novel sentences containing words the network has not previously seen. First, five sentences containing five instances of one new noun were exposed to the network for 50 sweeps, with the learning state of the network being captured every five sweeps. Over 50 sweeps, this translates to the new word being exposed between 13-14 times. Then, at the five sweep intervals, the network was tested for its ability to predict the new word in five previously unseen sentences, and the node activations of the new word was recorded. This was done for four

new nouns – one for each noun category. This was meant to be analogous to examining the ability of a child to name a new word in a cued context after hearing the word a certain number of times.

At the same time, the AP value was also determined for individual noun categories, to be able to compare improvement in AP with word learning across corpus sizes.

2.4.2 Results and Discussion

Figure 2-4 shows the average output unit prediction over training averaged over five new nouns that were taught to the networks after 140,000 sweeps. Higher node activation indicates better performance on the prediction task. Figure 2-5 shows the AP values of the network over time.

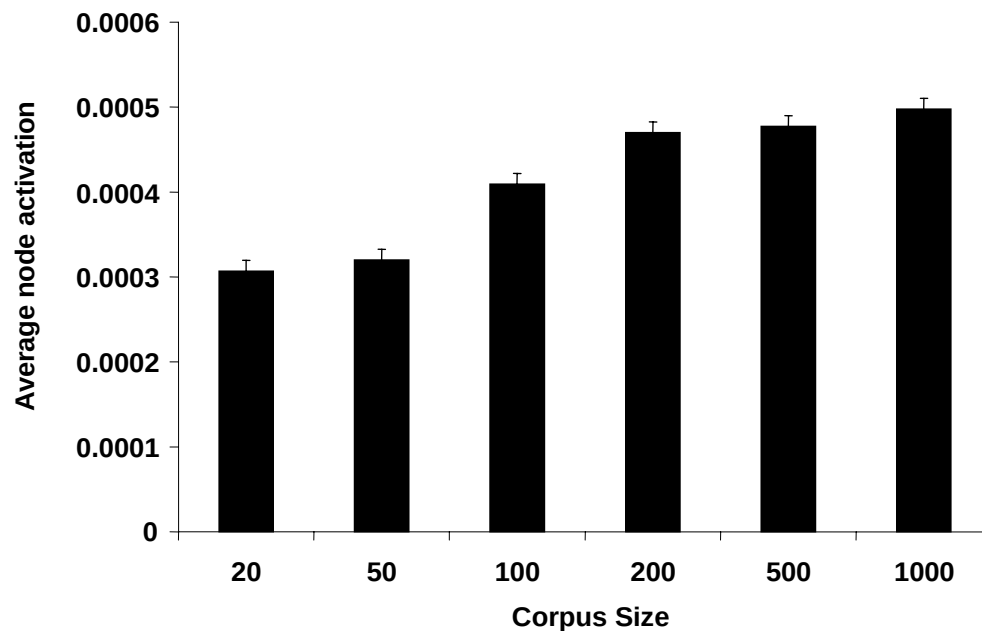


Figure 2-4. Word learning across size of input. Average node activation is plotted across corpus size for the prediction of the new word in a novel context.

From Figure 2-4 we see the trend that larger corpus sizes provided an earlier advantage in new word learning, with new word being predicted in appropriate contexts earlier in training than occurs in networks that have been exposed to smaller corpus sizes. Consistent with our hypothesis, there were significant differences as a function of size of training corpus $F(5, 234) = 42.4, p < 0.0001$. Post-hoc tests using Tukey's HSD (Table 2-1) revealed that word learning between 20 and 50 sentences were comparable, but smaller than all other training input sizes. Additionally, word learning between 200, 500, and 1000 sentences did not differ. However, the word learning with the 100 sentence corpus was larger than 20 and 50 sentences, but smaller than 200, 500 and 1000 sentence corpora.

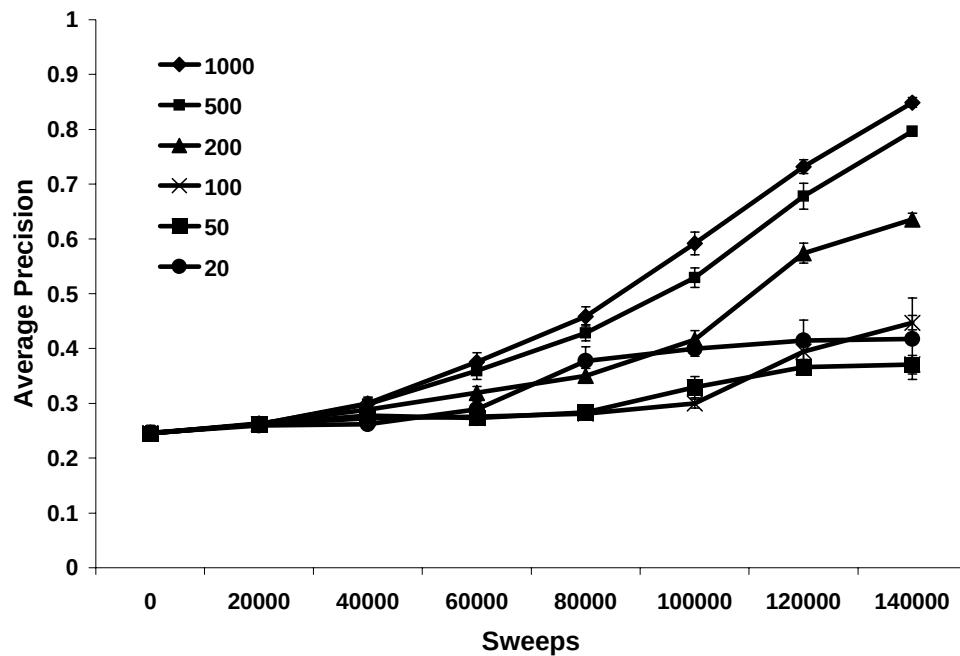


Figure 2-5. Change in average precision values over training for each corpus size.

Table 2-2. Word learning means and standard deviations for different training corpus sizes

Corpus size in sentences	Mean (SD)
20	.307 _a
50	.320 _a
100	.409 _b
200	.470 _c
500	.477 _c
1000	.498 _c

Note: Subscripts indicate post-hoc comparisons using Tukey's HSD. Means that do not share a common subscript are significantly different at $p < .05$

Next, AP values at the end of training were compared across size. It was predicted initially that higher AP values (which reflect greater category coherence) would also be associated with larger training corpora. Figure 2-5 shows the change in AP by corpus size over training. There were significant differences at the endpoint in training of AP value as a function of size of training corpus, $F(5, 54) = 404.21$, $p < 0.0001$. Post-hoc tests using Tukey's HSD (Table 2-3) revealed that AP differences were higher between larger corpus sizes, except between the 100 and 20 sentence corpus, where there was no difference. This supports the hypothesis that like word learning, higher AP values are attained by the network over time with networks that have the benefit of larger corpora.

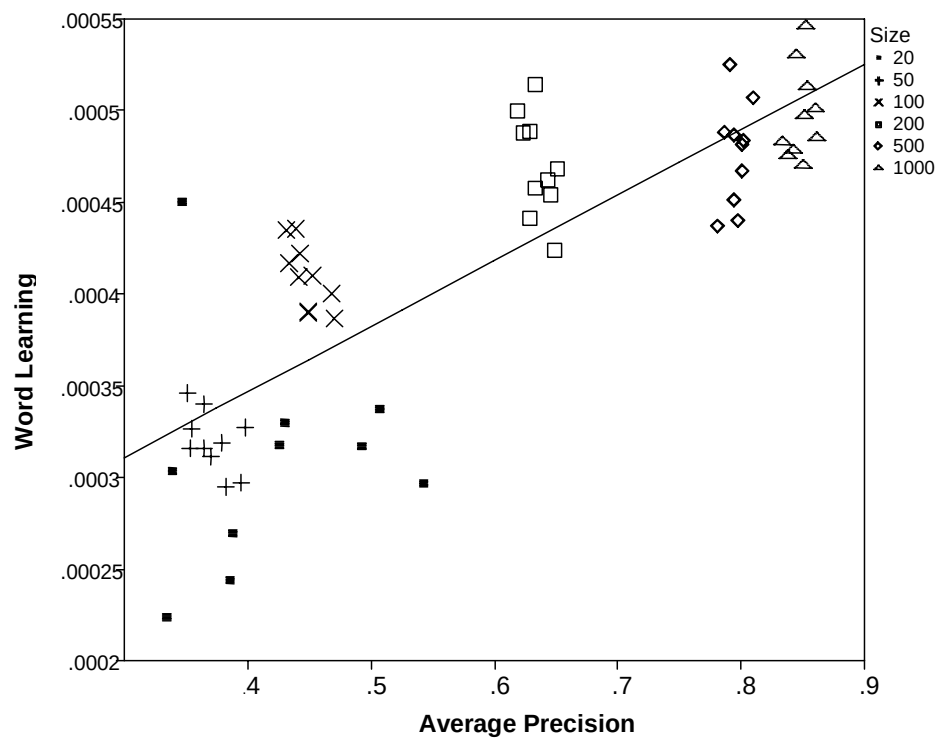


Figure 2-6. Average node activation across average precision value for each corpus size.

Table 2-3. Average precision means and standard deviations for different training corpus sizes

Corpus size in sentences	Mean (SD)
20	.418 _a (.074)
50	.371 _b (.017)
100	.447 _a (.0133)
200	.636 _c (.011)
500	.796 _d (.008)
1000	.849 _e (.009)

Note: Subscripts indicate post-hoc comparisons using Tukey's HSD. Means that do not share a common subscript are significantly different at $p < .05$

In order to better ascertain the relationship between word learning and category coherence, a simple regression was conducted with the AP values from the final point in training for each corpus size across the word learning node activation values. Figure 2-6 shows that higher AP scores were very highly correlated with better rates of word learning across corpus size, $R^2 = 0.68$, $F(1, 58) = 123.27$, $p < 0.0001$.

In sum, these results are consistent with previous findings in the acquisition literature that increased exposure to language results in better lexical acquisition skills. Here, networks that were exposed to input that was both more varied, and had more tokens reaped the benefits of being better able to learn new nouns that they were subsequently exposed to. Additionally, this simulation also supports the hypothesis that better noun learning is indeed associated with increased coherence of noun categories. This is consistent with both the notion that categorization abilities that appear around the time lexical learning improves could index better category structure, as well as the idea that it is possible to influence cognitive development (in the specific sense of inducing category structure) from linguistic input alone.

But can subtler differences in input also affect rate of word learning and category structure? It has long been known that the vocabulary to which children are exposed is skewed, in the sense that there are dramatic differences in the frequency of different lexical items (Bates, 1988; Broen, 1972). More recent work suggests that not only do these differences vary across families and children, but that the frequency with which different members of a category occur may play a role in learning and generalization. Bybee (1995) has noted such effects in the domain of morphological generalization, and Goldberg and colleagues present experimental findings that suggest that the presence of a high

frequency category exemplar during learning can facilitate the learning of categories (Casenhiser & Goldberg, 2005; Goldberg, Casenhiser, & Sethuraman, 2004). In the next study, we test this possibility by keeping the type frequency (number of different words) constant, but varying the token frequency of individual words.

2.5 Experiment 2: Effect of frequency

2.5.1 Methods

2.5.1.1 Input

The training input was constructed with the same characteristics from Experiment 1, with the same number of sentences, semantic and syntactic relations and vocabulary. The major difference was that the frequency of the word tokens was altered. There were two conditions: Even frequency and Uneven frequency. In the Even frequency condition, all members in a category were equally frequent, while in the Uneven frequency condition, one member of a category was much more frequent than the other members, while all other members shared the same low frequency.

2.5.1.2 Training

Each word frequency condition was presented to 10 SRNs with 50 hidden units and trained on a next-word prediction task. Learning rate was 0.01, no momentum was used, and training was carried out to 140,000 sweeps. Thus, the network saw an equal number of word types and each corpus the same number of times, but was exposed to different frequencies of the same words.

2.5.1.3 Analysis

Analysis proceeded in identical fashion as in Experiment 1, with AP values and new word learning being measured at regular intervals.

2.5.2 Results and Discussion

Figure 2-7 shows the AP values over training for the networks trained in each condition. There were no differences between AP values before and at 60,000 sweeps, nor at 140,000 sweeps. However, there were significant differences in AP values at 80,000 sweeps, $F(1, 144) = 920.77, p < 0.0001$, at 100,000 sweeps, $F(1, 144) = 926.57, p < 0.0001$, and at 120,000 sweeps, $F(1, 144) = 229.93, p < 0.0001$. These results suggest that there is an advantage for the network that sees words presented with uneven frequency, such that it enjoys an early boost that levels off, while the other network catches up. In other words, the token frequency manipulation did aid in initial category formation as predicted. (Note that, because of the limited size of the vocabulary, there is a ceiling effect in performance such that all learning regimes converge on the same level of performance. A similar phenomenon is observed with growth curves from the CDI: learning appears to level off for all children. But this is because the CDI tests a fixed (and relatively small) set of words. Eventually, like our networks, all children learn the words in this limited set. Therefore, such growth curves are most informative in the middle regions, above the floor and below the ceiling of performance.)

Next, the network was probed for new word learning at 80,000 and 100,000 sweeps, where the largest difference in AP values was found. The results of this learning at 80,000,

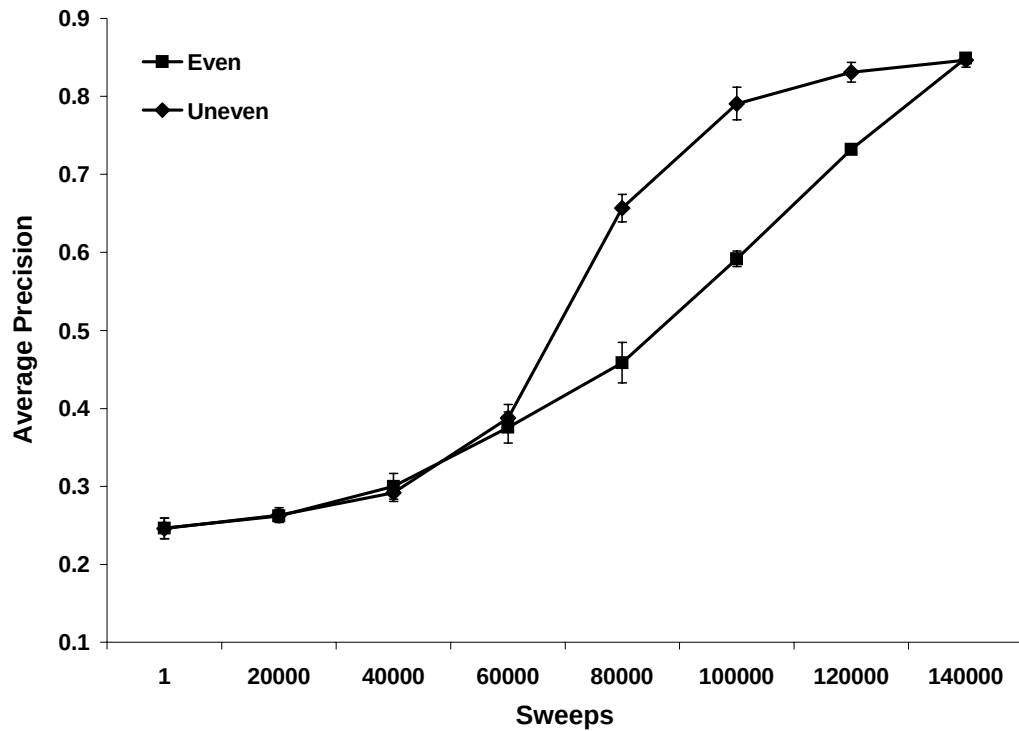


Figure 2-7. Average precision values over training and frequency condition.

100,000 and 140,000 sweeps can be seen in Figure 2-8. When measuring the rate of new word learning by taking the average node activations across the 50 sweeps of training, new noun learning for the uneven condition shows an advantage at 80,000 sweeps, $F(1, 312) = 34.07$, $p < 0.0001$, but not at 100,000 sweeps, $F(1, 312) = 0.03$, ns . The result for 80,000 sweeps but not 100,000 falls in line with predictions that uneven token frequency in each word category should provide some sort of benefit in word learning that is tied to improvements in categorical structure. In order to examine word learning when there are no differences in AP values, Figure 2-8 also plots how well new nouns are being learned at the end of training at 140,000 sweeps, where no differences in AP values were observed. Examining this portion of the graph, the networks trained with an evenly distributed word

frequency have a significant learning advantage, $F(1, 312) = 5.51, p < 0.01$. Although there are no differences in AP value at this point, we do observe a difference in word learning in the Even condition. This seems to reverse the trend seen at 80,000 sweeps in training where the uneven condition held the advantage in both word learning rates and AP values. It appears that as differences in AP values get smaller between the two frequency conditions, word learning in the Even condition improves.

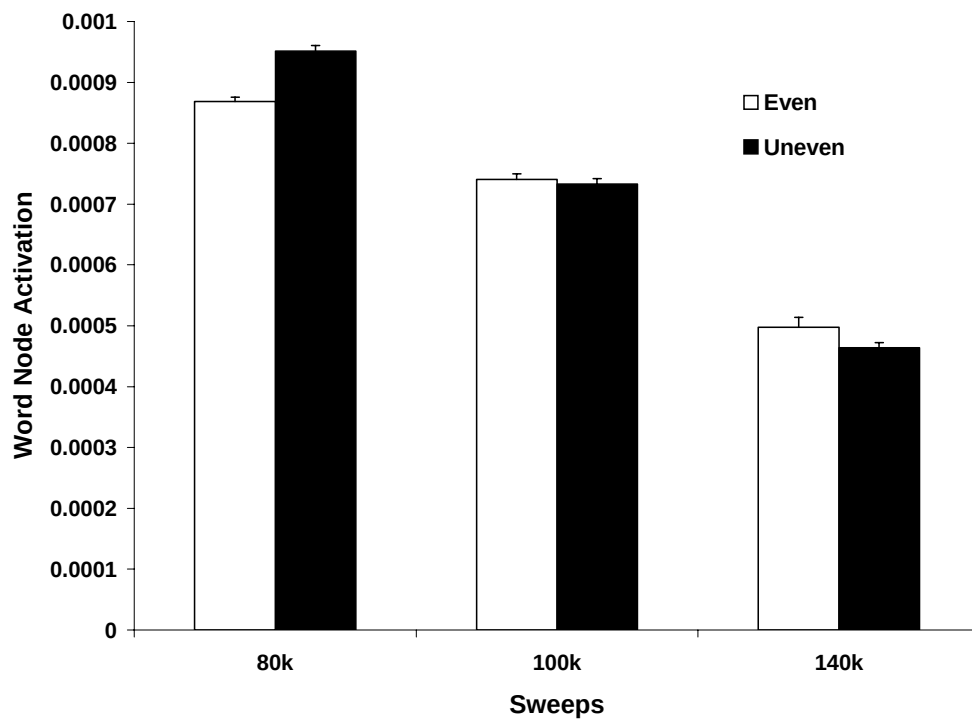


Figure 2-8. Rate of word learning at 80k, 100k and 140k sweeps in both frequency conditions as measured by average word node activations over training.

In order to examine more closely how differences in AP values between the two conditions relate to differences in word learning rates, normalized differences of AP scores and word learning between networks with the same initial random weight setting (this is

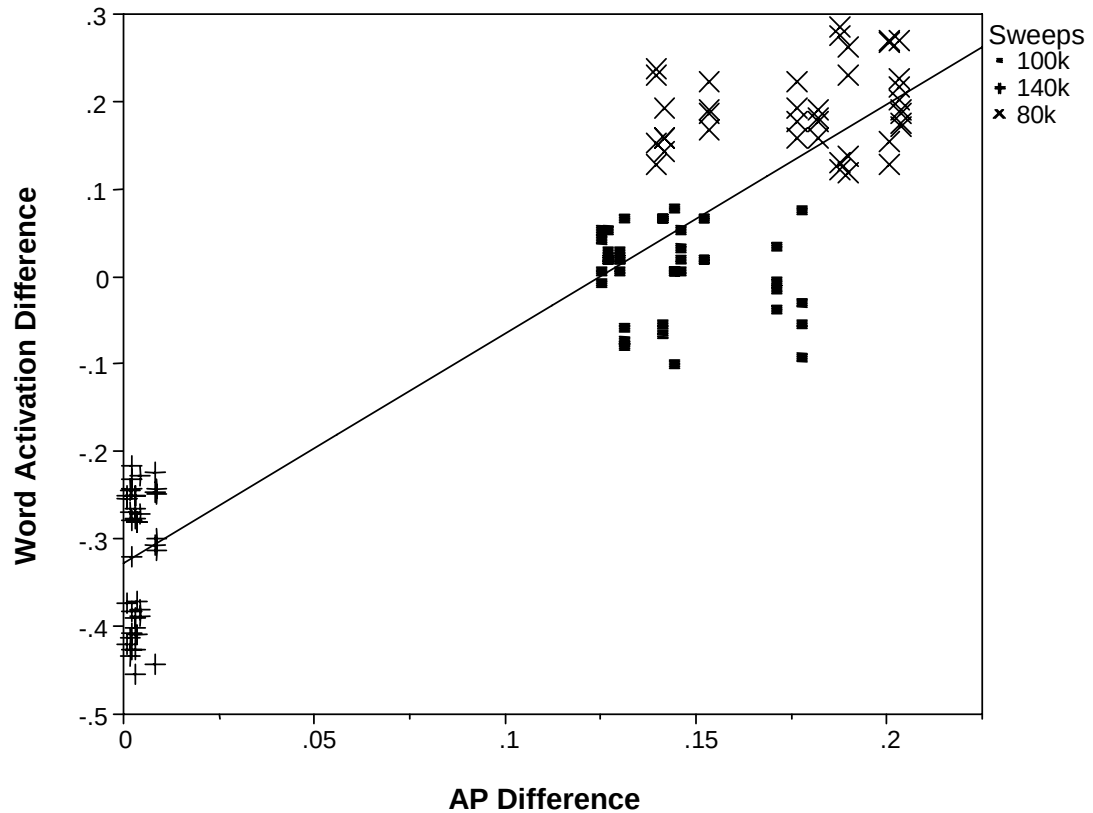


Figure 2-9. Differences between even and uneven new word node activations plotted across difference in average precision values.

analogous to using the same human subject) at 80,000, 100,000 and 140,000 sweeps are plotted in Figure 2-9. Simple linear regression reveals a highly significant relationship between differences in AP and differences in word learning between the even and uneven condition $R^2 = 0.85$, $F(1, 28) = 703.58$, $p < 0.0001$. This analysis suggests that even though we find differences between the two word frequency conditions in rates of new word learning at 140,000 sweeps when there are not differences in AP scores, and that we find no differences at 100,000 sweeps, even though there are differences in AP scores, there is still a very predictable relationship between changes in AP values and new word learning. This indicates

that the overall relationship between the difference in learning rate and coherence in each condition, even though raw values might fluctuate. Here, our regression model reveals a very highly significant and positive relationship between differences in AP and subsequent differences in new word learning, such that when conditions improve new word learning ability in a particular network, we can also expect improved AP scores.

Finally, we turn to a third way in which input conditions might differ and ask what the effect of grammatical complexity might be on new word learning. The literature on this point reports mixed findings, perhaps because different measures of grammatical complexity are used in different studies. Recognizing that there are many dimensions along which such complexity might be defined, we begin with a very straightforward manipulation. We will define complexity in terms of the number of different arguments that are involved in a construction. The grammatically simple condition will involve only transitive and intransitive constructions (NVN and NV); the grammatically complex condition will additionally include ditransitive and sentential complement constructions (NVNN, NVNV, NVNVN).

2.6 Experiment 3: Effect of grammatical complexity

2.6.1 Method

2.6.1.1 Input

Two 1000 sentence corpora were constructed, one for the Simple condition and one for the Complex condition. The Simple corpus was constructed with the same characteristics as the 1000 sentence corpus in Experiment 1. The Complex corpus was constructed with 10 additional verbs that belonged to two new verb categories: PSYCH (5)

and TRANSFER (5). Semantic and syntactic relations between these two verb categories are included in Appendix 2. Examples of sentences containing verbs in these categories are also included in Appendix 3. No new nouns were added. Thus, the makeup of tokens in the complex grammar contained 95 total words (52 nouns and 43 verbs; the network architecture was adjusted to reflect the larger input and output vectors).

The new verb categories allowed for additionally syntactic complexity by allowing for three more complex constructions to be added to the already present NV and NVN constructions. TRANSFER verbs allowed for ditransitive constructions of the form: NVNN. PSYCH verbs allowed for NVNV or NVNVN) constructions, with PSYCH verbs only occurring in the first verb position in these sentences.

2.6.1.2 Training

Each of the two corpora were presented to 10 SRNs with 50 hidden units and trained on a next-word prediction task. Learning rate was 0.01, no momentum was used, and training was carried out to 140,000 sweeps. Thus, the network saw an equal number of word tokens. Because the average length of sentences was different between each condition, each corpus was not seen the same number of times.

2.6.1.3 Analysis

Analysis proceeded in identical fashion as in Experiments 1 and 2, with AP values and new word learning being measured at regular intervals.

2.6.2 Results and Discussion

Figure 2-10 shows the AP values over training in networks with the syntactically simple and complex input. The graph shows that the simple input has significantly higher AP values at 80,000 sweeps, $F(1, 18) = 195.02$, $p < 0.0001$, 100,000 sweeps, $F(1,18) = 674.05$, $p < 0.0001$, 120,000 sweeps, $F(1,18) = 628.42$, $p < 0.0001$, and 140,000 sweeps, $F(1,18) = 637.84$, $p < 0.0001$. AP is equivalent between each syntactic condition earlier. These findings suggest that simpler grammatical constructions do indeed aid in early categorical formation, because simpler syntax in this manipulation displayed higher AP values after training.

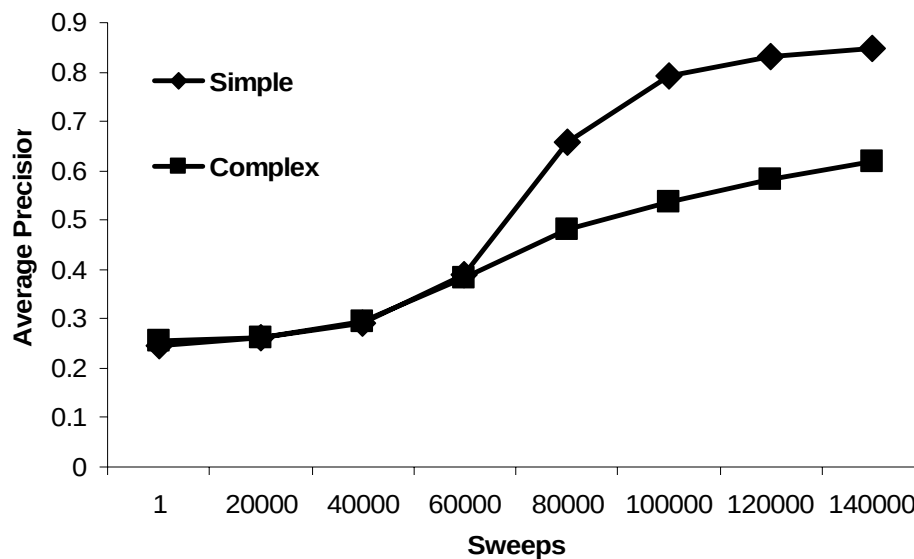


Figure 2-10. Average precision values across training in simple and complex input conditions

Figure 2-11 relates these findings to new word learning. This graph shows how well the networks trained with each kind of input learned new words after 140,000 sweeps. Here, the networks trained with simpler input indicate stronger activation to predict new

nouns. The rate of new learning was then assessed by taking the average activation value of the word node to predict the new noun from the initial value. We find that there is also a significant difference between new word learning with simple syntax learning showing higher rates of word learning than more complex syntax, $F(1, 78) = 130.41$, $p < 0.0001$. These results are consistent both with the hypothesis that simpler grammar should improve lexical acquisition, and that higher category coherence also predicts better word learning.

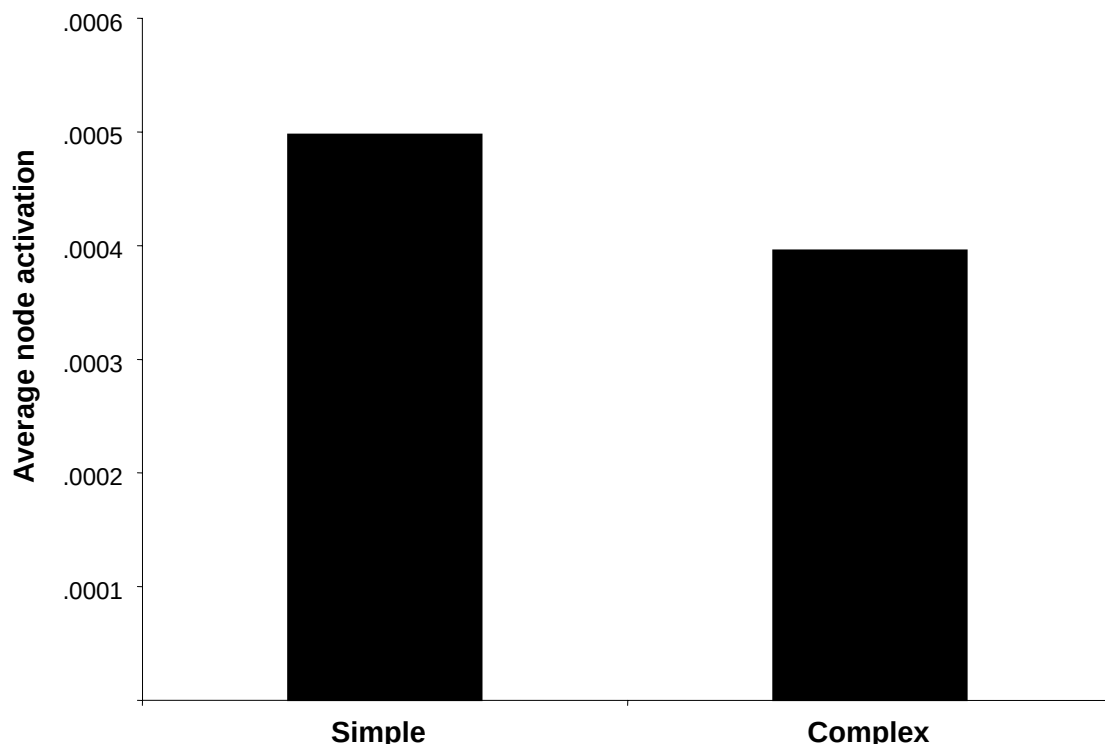


Figure 2-11. Rate of word learning across complexity of input.

2.7 General Discussion

These three experiments tested two predictions that follow from the hypothesis that categorization is used as a tool in lexical acquisition.

- 1) Variation in language input that affects lexical acquisition also affects development of category coherence in a similar manner; and
- 2) It is the development of categories themselves, and not solely language input, that facilitates lexical acquisition.

The results of these three experiments support these predictions. All three show direct relationships between input condition and corresponding improvements in lexical acquisition and category coherence.

A very clear illustration of this relationship is demonstrated in Figure 2-6, from Experiment 1, where increasing amounts of input positively influenced both category coherence and word learning. This supports the idea that there is a relationship between category development and speed of word learning. However, it is still not clear from these results if this relationship is of the nature as described in our second prediction, or is driven by language input driving both factors independently, but in coincidentally similar ways.

A key finding that addresses this issue comes from Experiment 2, where the frequency of the words in each category was manipulated. Here, while the nature of the input remained constant—in terms of the richness of the vocabulary—we see a direct relationship between improvements in category and lexical development. This suggests that lexical acquisition is affected by changes that the input has on category structure.

2.7.1 Word learning mechanisms

Much of the work that has been devoted to explaining word learning focuses on the dramatic changes in rate of lexical acquisition that occur during the middle of the

second year of life. Two basic types of theories have been proposed. One class of theories hypothesizes that language specific constraints and principles appear during the vocabulary spurt and have the effect of improving the ability of children to acquire new words. These accounts have tended to emphasize the language-specific nature of the constraints involved in word learning (Markman, 1989; Woodward & Markman, 1998), although it is also possible that these domain-specific word learning constraints may arise from more domain-general processes, as in the ‘lexical principles’ model (Golinkoff, Mervis, & Hirsh-Pasek, 1994).

An alternative to the constraints and principles view is one that words are learned through domain-general processes that make note of statistical associations between words and other properties that accompany their referents. These theories are entirely compatible with a view that other cognitive processes like categorization may aid in knowledge generalization from similar words in the same category. The simulations we report here illustrate how such a mechanism might work. The simulations demonstrate that a single learning mechanism that does not change over the course of development can account for a number of ways in which input differences seem able to influence word learning ability. The effects are mediated by category knowledge; interestingly, the development of these categories can be manipulated not only by quantitative variations in the input, but also by differences in grammatical complexity and frequency distributions across the input vocabulary.

These results are thus consistent with Gopnik & Meltzoff’s (1993; , 1997) account of the use of categorization as a tool in learning language as a ‘complex bi-directional interaction’. We find a direct inter-relationship between improvement in category scores

and word learning, when starting from clean conceptual slate. By training neural networks, we were able to examine how linguistic input might serve to carve out categorical space in the absence of any other kind of perceptual input, and how this categorical space in turn improves proficiency in lexical acquisition. In this way, we find support for a bidirectional influence of language and categorical development. Through semantic information that is encoded solely in language, it is possible to find relationships between improvements in linguistic ability and category development in the absence of perceptual input.

Overall, we have found that a single domain-general mechanism can account for a number of patterns in child word learning. Initially, our networks show very low category coherence, as it learns about individual items. This pattern is also observed in children, where analysis of early vocabulary shows that children tend to learn about basic level words before superordinate or subordinate items (Mervis, 1983). Category coherence improves because the neural networks eventually learn to categorize word items that are more similar to each other. More input aids in this process, by allowing the network to have a larger variety of experience and examples in which it may use to more appropriately classify items into groups. Simpler syntax is useful in allowing the network to process simpler relations that are easier to understand and more easily grouped. Also, having a highly frequent exemplar in each noun category was useful to allow the networks to understand one example very well in a variety of contexts, thereby allowing other members of the same category to be organized more easily into a group. Word learning was improved in cases of better developed category coherence, because the networks were able to generalize from its knowledge of other members in the category to new words seen in similar contexts.

2.7.2 Role of input

The role of language input is highly emphasized in this paper. However, it is important to acknowledge several ways in which the experience of the networks differs from that of children.

First, in these simulations we are limited to using a simplified artificial language that does not represent the full richness and complexity of natural language. Undoubtedly there are many other factors in CDS that may affect the outcome of this study. For instance, earlier disagreement about the role of structural complexity in language might actually be related to a connection in increasing complexity of CDS with age. Hoff & Naigles (2002) found that increased syntactic complexity in CDS was a better predictor of vocabulary size in 2;0 year olds. On the other hand, Brent & Siskind's (2001) study suggest that simpler complexity is beneficial was achieved with 0;9 to 1;3 month olds. It could be that increased complexity is beneficial for older children, like those in Hoff & Naigles (2002). A similar sort of result is reported by Elman (1993), where networks trained on input complexity that was incrementally increased in complexity demonstrated better learning than starting off with complex input initially.

Second, the input used in this study completely ignores the role that other social and perceptual cues may play in language learning. We have no doubt that input in other forms and modalities is crucial in both language and cognitive development. In fact, there is evidence that around the time of the vocabulary spurt, children seem to be making important developments in their social abilities as well. Thus, we emphasize that we do not believe that information from verbal input is the sole determinant of conceptual development.

Nonetheless, it is a striking result that the linguistic input alone—absent the experiential input available to real children—is such a rich source of information about category structure. Furthermore, we suspect that the additional information available to children will be learned through a similar kind of association mechanism that organizes experience based upon similarity. Essentially, input comes in many forms, but the underlying principles for learning is the same for all of them.

2.8 Conclusions

It is clear from these experiments that there are potentially many factors in CDS that may influence both a child's ability to learn words and her categorical knowledge. Here it is implied that disadvantages that arise from deficient input will persist due to influences that remain from the cognitive implications of this deficit. However, by identifying at least one underlying cognitive factor that is affected by change in language input, it is possible that interventions may be designed for children in normally impoverished linguistic environments. For instance, Hart & Risley (1995) mention somewhat pessimistically that differences in the amount of language input that are related to socioeconomic status are so large that it would require a continuous intervention of 40 hours per week to make up for 'lost input'. Perhaps focused training on category development may boost word learning ability in these children that could at least partially make up for deficiencies in language experience by aiding them to make the most efficient use of language that they do hear. This may also apply to cases where children may have difficulty learning language, or where there may be a delay of input due to conditions like deafness. Indeed studies by Smith, Jones, Landau, Gershkoff-Stowe & Samuelson (2002)

have shown that training children to attend to shape rather than texture of objects boosts the number of object words known several months later. It is still left to be determined if this kind of approach could also apply to other types of categories, and other aspect of input. There are all important questions that merit further study.

In conclusion, while there are a number of limitations to this study, both in its simplification of language input and in the exclusion of other forms of perceptual input, the results we have found still are able to account for several patterns in language acquisition in children. Indeed, our results do mimic trends that we have already seen in the developmental literature. For instance, we replicated findings by (Huttenlocher et al, 1991) and (Hart & Risley, 1995), that less exposure to linguistic input seems to put children at a disadvantage in language knowledge and learning. If anything, this type of replication suggests that the kinds of phenomena we have measured in our simulations, should also relate to learning in children. Most importantly, overall, this work provides a clearer understanding of a possible mechanism by which the development of category knowledge and word learning may be related.

Chapter 2, in full, is a reprint of the material as it appears in Borovsky, Arielle & Elman, Jeff. (2006). Language input and semantic categories: A relation between cognition and early word learning. *Journal of Child Language*, 33, 759-90. The dissertation author was the primary investigator and author of this paper. Permission to reprint was granted by Cambridge Journals, Cambridge, UK.

CHAPTER 3

LEARNING TO USE WORDS

3.1 Abstract

Humans have an amazing capacity to learn words from a single instance. Yet little is known about the nature of word knowledge that is rapidly acquired and how this is modulated by learning context. Adults read known and unknown words in strongly or weakly constraining sentence contexts and assessed word usage knowledge via subsequent plausibility ratings of these words as transitive verb objects while recording event-related brain potentials. N400 plausibility effects at the verb appeared when expected upcoming unknown words originally appeared in a strongly constraining context. These results demonstrate that one-shot word learning is modulated by contextual constraint and reveals a rapid mental process that incorporates information about the proper usage of novel words into the mental lexicon.

3.2 Introduction

Humans have an amazing capacity to learn many thousands of words – an ability unmatched by any other species. This ability enables children to form the building blocks of grammar (Bassano, 2000; Bates & Goodman, 1997; Caselli, Casadio, & Bates, 1999; Tomasello, 2000), and allows humanity to incorporate new ideas into culture. Partly in recognition of the foundational nature of word learning in human cognition, an enormous body of research on word learning in young children and adult bilingualism has accumulated over the past fifty years. In contrast, relatively little is known about word

learning (lexical acquisition) by adults in their native language, despite the fact that most words are acquired after early childhood. Estimates of adult vocabulary levels vary widely, but generally fall between 40000–150000, whereas a pre-literate five or six year old will know only 2500-13000 words (Aitchinson, 1994; Beck & McKeown, 1991; Bloom, 2000; Pinker, 1994). Moreover, adults and older children learn words differently from younger children. Pre-literate children often learn new words through explicit naming and reference, while school age children and adults acquire words almost entirely incidentally in various language contexts, especially reading (Jenkins, Stein, & Wysocki, 1984; Nagy, Anderson, & Herman, 1987; Nagy, Herman, & Anderson, 1985; Sternberg, 1987). This mode of learning can be remarkably fast. Under the right conditions, only a single exposure to a novel word may be sufficient for a learner to infer its probable meaning (Carey & Bartlett, 1978; Dollaghan, 1985; Heibeck & Markman, 1987).

The goal of this research was to understand what those conditions might be, while focusing on the importance of sentence context. We explored this issue by measuring event-related potentials (ERPs) as adults read known words and unknown pseudowords in sentence contexts that strongly or weakly constrained word meanings.

For example, participants may have read a novel word “marf” in either a weakly constraining context such as: “She walked across the room to Mike’s messy desk to return his marf,” or in a more strongly constraining context like: “He tried to put the pieces of the broken plate back together with marf”. (Responses to unknown words were compared to control conditions where known words such as “glue” appeared.) Although both contexts provide information about the meaning and proper contextual usage of “marf”, intuitively, a more specific understanding of its usage could be inferred in a stronger constraint

context. We tested this hypothesis by asking participants to rate the plausibility (i.e., acceptability of the usage of this word) in two simple test sentences presented immediately after the initial context sentence: e.g., “They used the marf.” and “She drove the marf”. These sentences always followed the form “Pronoun–Transitive Verb–Article/Pronoun–Target word”. Since the test sentences appeared immediately after the context sentences, the target word of each test sentence (in this case marf) was completely predictable. Consequently, all of the information needed to determine if the (known or unknown) word was used appropriately was available at the verb (in this case, the plausible “used” or the implausible “drove”). Thus, our interest was in the processing that occurred at the verb.

By measuring the amplitude of the N400 component of the ERP to the verb preceding the upcoming target word object in this task we can determine whether or not a reader has gleaned enough meaning from a novel word to determine if it was appropriately used in a new relatively sparse sentence context – i.e., whether or not the target word could legitimately serve as an object of that verb. The N400 is a negative going brainwave from 250-500ms (peaking around 400 ms) after the presentation of any potentially meaningful stimulus (Kutas & Hillyard, 1980). N400 amplitudes are larger when a word is unknown, used inappropriately, is less frequent, or is a pseudoword (Bentin, 1987; Kutas & Hillyard, 1980; McLaughlin, Osterhout, & Kim, 2004; Mestres-Misse, Rodriguez-Fornells, & Munte, 2006; Perfetti, Wlotko, & Hart, 2005). In addition the N400 has recently been identified as a component sensitive to word knowledge and learning in both children (Friedrich & Friederici, 2004; Friedrich & Friederici, 2005a, 2005b, 2006; Mills, Coffey-Corina, & Neville, 1997; Torkildsen et al., 2006; Torkildsen, Syversen, Simonsen,

Moen, & Lindgren, 2007) and adults (McLaughlin, Osterhout, & Kim, 2004; Mestres-Misse, Rodriguez-Fornells, & Munte, 2006; Ojima, Nakata, & Kakigi, 2005; Perfetti, Wlotko, & Hart, 2005; Stein et al., 2006). By looking at plausibility effects on the preceding verb rather than the repeated novel word, we avoid known confounds due to word repetition: N400 amplitudes to words are reduced by repetition. Moreover, we can infer if knowledge of word usage is rapidly acquired by the extent that the N400 is modulated by the appropriateness of subsequent use of the word.

3.3 Methods

3.3.1 Participants

26 right-handed college students (9 M, 17 F) between the ages of 18-25 (mean 19.8) were given credit or paid \$7/hr for their participation. All participants were native English speakers, with no significant exposure to another language before the age of 12. All participants had normal hearing and normal (or corrected to normal) vision and reported no history of mental illness, learning disability, language impairment, drug abuse, or neurological trauma. An additional 11 participated but were not analyzed: 4 had excessive blinking or motion artifact, 3 because of equipment failure or experimenter error, and 4 reported a characteristic which disqualified them from analysis (3 had significant second language exposure as a child, 1 reported significant illicit drug use.)

3.3.2 Stimuli

Two types of sentences were selected for the study: 1) Context sentences, which provide High and Low constraint contexts for Known and Unknown word targets; and 2)

Test sentences, which were very short sentences, following the form “Pronoun–Transitive Verb–Article/Pronoun–Target noun”.

Eighty strongly constraining and 80 weakly constraining sentence fragments were selected from Federmeier and Kutas (2005). Each strongly constraining sentence fragment was completed by the word that obtaining the highest cloze probability, and these same 80 words were then paired with the weakly constraining contexts to yield plausible, low cloze probability endings. Each sentence pair was assigned one of 80 pronounceable nonwords to serve as an alternate sentence ending. This arrangement yielded four main Context sentence conditions with 40 sentences each: 1) High constraint sentences with Known word endings (High/Known), 2) High constraint sentences with Unknown word endings (High/Unknown), 3) Low constraint sentences with Known word endings (Low/Known), and 4) Low constraint sentences with Unknown word endings (Low/Unknown). Table 3-1 presents examples of each condition. Across all versions of the experiment, sentence-final target words in the two constraint conditions were counterbalanced, such that the same sentence and target ending did not appear twice in any version, but all possible combinations of sentence / target word pairs appeared in all versions. The purpose of this was to ensure that any differences in performance due to the properties of the words were balanced out across conditions.

Four test sentences were created for each high/low constraint sentence pair for use in a plausibility judgment task. Two of these sentences presented an implausible usage of the target word, and two plausible. Plausibility of the Test sentences was confirmed in a plausibility rating study conducted with different set of participants with each sentence and target word combination. All four sentences were used, and the verbs in each test

sentence were balanced so that they were paired with both a plausible and implausible usage of the target word (see examples in Table 3-1).

Table 3-1. Sentence examples in each condition

<u>Context Sentences</u>	
High/Known:	She walked across the room to Mike's messy desk to return his GLUE.
High/Unknown:	She walked across the room to Mike's messy desk to return his MARF.
Low/Known:	He tried to put the pieces of the broken plate back together with GLUE.
Low/Unknown:	He tried to put the pieces of the broken plate back together with MARF.
<u>Test Sentences</u>	
Plausible:	They used the GLUE / MARF.
(Known/Unknown)	He needed the GLUE / MARF.
Implausible:	She drove the GLUE/MARF.
(Known/Unknown)	He greeted the GLUE/MARF.

3.3.3 Neuropsychological measures

In addition to the experimental task, participants also completed a number of background neuropsychological tasks. These included a forward and backward digit span task, reading comprehension task, and Peabody Picture Vocabulary Task (PPVT).

3.3.4 Procedure

Subjects participated in a single experimental session conducted in a soundproof, electrically-shielded chamber, seated in a comfortable chair in front of a CRT monitor. Participants were asked to read the context sentences for comprehension even when

“nonsense/unknown” words appeared. Immediately following each context sentence, participants rated the plausibility of two subsequent sentences that ended in the exact same target word as the preceding context sentence. Participants were not instructed to generate an explicit meaning for Unknown words, and the rapid pace of the plausibility task was designed to minimize explicit naming strategies.

Each sentence presentation was preceded by a series of crosses (500 ms duration with a stimulus-onset-asynchrony varying randomly between 300-800 ms) to orient the participant toward the center of the screen. Sentences were presented one word at a time, and all but the final word appeared for 200 ms with a stimulus-onset-asynchrony (SOA) of 500 ms. Participants were asked to minimize blinking and movement as much as possible during sentence presentation. The final, target word appeared on the screen for 1400 ms, and was immediately followed by two plausibility judgment sentences. Each plausibility sentence was preceded by a series of question marks (400ms duration with randomly varied SOA of 100-300ms). Plausibility sentences were presented with identical timing to Context sentences. Participants were asked to make plausibility rating as soon as possible after each test sentence’s final word. Participants were compensated and debriefed at the conclusion of the session.

3.3.5 Recording

Scalp potentials were continuously recorded from 26 geodesically arranged sites using an ElectroCap with tin electrodes, referenced to the left mastoid. Potentials were digitized at a sampling rate of 250 Hz and bandpass filter of 0.1-100Hz with Grass Amplifiers. Impedances were kept below 5 k Ω .

3.3.6 Data Analysis

Data were re-referenced offline to an average mastoid. Trials contaminated by eye movements, blinks, excessive muscle activity, or amplifier blocking were rejected off-line before averaging. ERPs were computed for epochs extending from 100 ms before stimulus onset to 920 ms after stimulus onset. Averages of artifact-free ERP trials were computed for the target words in the four learning conditions (High/Known, High/Unknown, Low/Known, Low/Unknown) as well as to verbs in all four plausibility test conditions (Plausible/Known word, Plausible/Unknown word, Implausible/Known word, Implausible/Unknown word) after subtraction of the 100 ms pre-stimulus baseline.

3.4 Results and Discussion

3.4.1 Accuracy

During the task, behavioral and brainwave responses were simultaneously collected from participants, who were asked to decide as quickly and accurately as possible if the sentence final target word was being used appropriately in each of two test sentences. Plausibility accuracy ranged between 69% and 90% (Table 3-2). A three-factor ANOVA with factors of Constraint (Low and High), Word Type (Known and Unknown) and Plausibility (Plausible and Implausible) revealed effects of Constraint [$F(1,200)=25.4$, $p<0.0001$], with reduced accuracies for Low constraint items, and Word Type [$F(1,200)=45.6$, $p<0.0001$], with increased accuracy to Known words. There was no main effect of Plausibility [$F(1,200)=1.79$, $p=.18$]. There was a Constraint x Type interaction [$F(1,200)=5.76$, $p=0.0173$] and Type x Plausibility interaction [$F(1,200)=15.82$, $p<0.0001$]. Tukey analyses revealed that Known word accuracy was unaffected by Constraint whereas

accuracies for Unknown words were reduced in Low constraint contexts. Together this suggests that participants had a more difficult time understanding the meanings of Unknown words, with the greatest difficulty appearing in the Unknown/Low condition. These results were identical when conducted on arcsin transformed percentages, suggesting these effects were not driven by Known word ceiling effects.

Table 3-2. Accuracy of responses to plausibility task

Condition	% Accuracy (SD)
<u>High/Known</u>	
Plausible	90 (2.84)
Implausible	88 (3.20)
Overall	89 (3.02)
<u>High/Unknown</u>	
Plausible	80 (4.09)
Implausible	84 (3.84)
Overall	82 (4.02)
<u>Low/Known</u>	
Plausible	89 (2.61)
Implausible	81 (3.52)
Overall	85 (3.55)
<u>Low/Unknown</u>	
Plausible	69 (3.12)
Implausible	75 (4.54)
Overall	72 (4.07)

3.4.2 Context sentence ERPs

ERP brain potentials to the final words of the context sentences are shown in Figure 3-1. N400 mean amplitude was measured between 300-500ms post final word onset at four centro-parietal electrode sites (RMCE, LMCE, MiCE, MiPa) where N400 effects are typically largest. A two-factor repeated measures ANOVA with factors of Word Type (Known and Unknown) and Constraint (High and Low) revealed main effects of Constraint [$F(1,25)=24.92$ $p<0.0001$], Type [$F(1,25)=22.44$, $p<0.0001$], as well as a Constraint x Type interaction [$F(1,25)=23.45$, $p<0.0001$]. Tukey tests indicated that these effects were driven by Known/High target endings being significantly more positive than every other condition, with no other condition differing from any other. Thus ERP responses to Known words show a modulation by sentential constraint. By contrast, Unknown words do not show any effect of constraint in the N400 region. At first glance, this would suggest that the initial processing of Unknown words did not differ by constraint.

Additionally, an extended late positive component (LPC) was elicited for both Known and Unknown words. A two-factor repeated measures ANOVA with factors of Word Type (Known and Unknown) and Constraint (High and Low) also was conducted on the LPC mean amplitude between 500-700 ms post final word onset at the same recording sites. This analysis revealed effects of Constraint [$F(1,25)=8.71$ $p=0.0068$], with High (relative to Low) constraint target words showing greater positivity, and a main effect of Type [$F(1,25)=14.34$ $p=0.0009$], with Known (relative to Unknown) words showing greater overall positivity in this time period. A Constraint x Type interaction was also found [$F(1,25)=14.18$, $p=0.0009$], which Tukey tests revealed reflected Known/High targets being significantly more positive than any other condition, which did not differ from each other.

Thus, there was no reliable signature in the ERP to suggest that Unknown words in context sentences had been differentially processed as a function of contextual constraint.

3.4.3 Plausibility ERP effects

ERPs to verbs in Plausible and Implausible Test sentences are shown in Figure 3-2. Note that it is at this point in the sentence where all the information needed for a plausibility judgment first becomes available. N400 amplitude differences in plausibility were measured between 300-500ms at four centrally located electrodes (RMCE, LMCE, MiCe, MiPa) and subjected to a three-factor repeated measures ANOVA with factors of Word Type (Known and Unknown) and Constraint (High and Low) and Plausibility (Plausible and Implausible). This analysis revealed no effect of Constraint ($F < 1$), or Type [$F(1,25)=1.2984$, $p=0.2653$]. A significant effect of Plausibility [$F(1,25)=46.5553$, $p<0.0001$] was driven by more positive amplitudes to Plausible verbs than Implausible verbs. No significant two-way interactions were observed but a three-way interaction of Constraint x Type x Plausibility was significant [$F(1,25)=5.1258$, $p<0.0325$]. Posthoc Tukey tests revealed plausibility effects in N400 amplitude in all conditions, except for where verbs preceded Unknown/Low words. These results suggest that adults do have significant understanding of novel word meanings after a single presentation, but only when the unknown word initially appeared in a strongly constraining context.

In order to examine the relationship between the plausibility ERP effects observed in this study and the neuropsychological measures, we correlated the N400 difference between plausible and implausible verb usages in the Unknown/High, Unknown/Low, Known/High and Known/Low conditions with PPVT, reading comprehension, and digit span. R-square

correlation values and significances are shown in Table 3-3. Correlations between neuropsychological measures and N400 plausibility effects in all experimental conditions. We did not find significant relationships between the plausibility effects and vocabulary knowledge or digit span. However, there were significant relationships between our reading comprehension measures and plausibility effects – but only for novel word conditions. This result suggests that our task may tap into the same cognitive mechanisms utilized during our reading comprehension task, which required participants to quickly integrate, remember, and infer meaning from short paragraphs. This is not unlike the task presented before our participants, who needed to quickly integrate novel information about word meanings from sentence contexts. In addition, even though we did not see a significant effect of plausibility for the Unknown/Low constraint condition, the reading comprehension scores were still able to explain significant amount of variance on this task, with better reading comprehension scores being associated with larger plausibility effects. This suggests that individuals who are more skilled at reading comprehension tasks may be able to learn significant information about novel word meanings in both high and low constraint contexts.

Table 3-3. Correlations between neuropsychological measures and N400 plausibility effects in all experimental conditions

	PPVT	Digit Span	Reading Comprehension
<u>Condition:</u>			
Known / High	0.031	0.003	0.037
Known / Low	0.068	0.138	0.094
Unknown / High	0.089	0.018	0.273*
Unknown / Low	0.049	0.005	0.310*

Note: * - indicates p value < 0.03

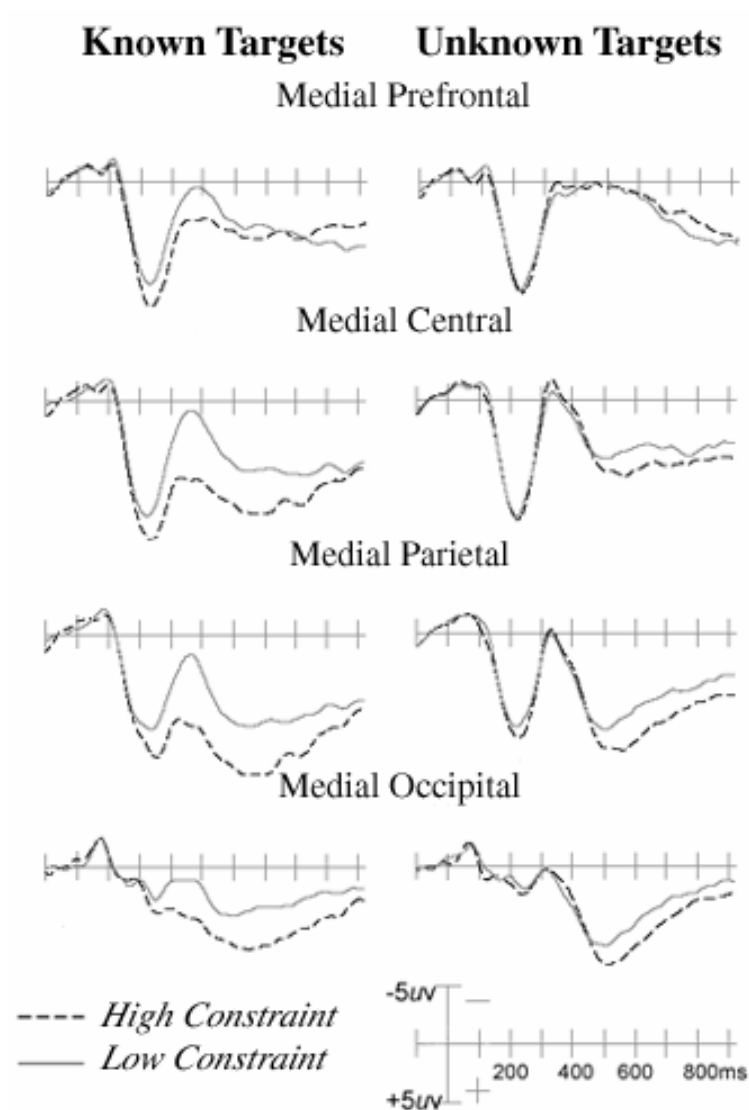


Figure 3-1. ERPs to Known and Unknown target words in context sentences at midline electrode sites

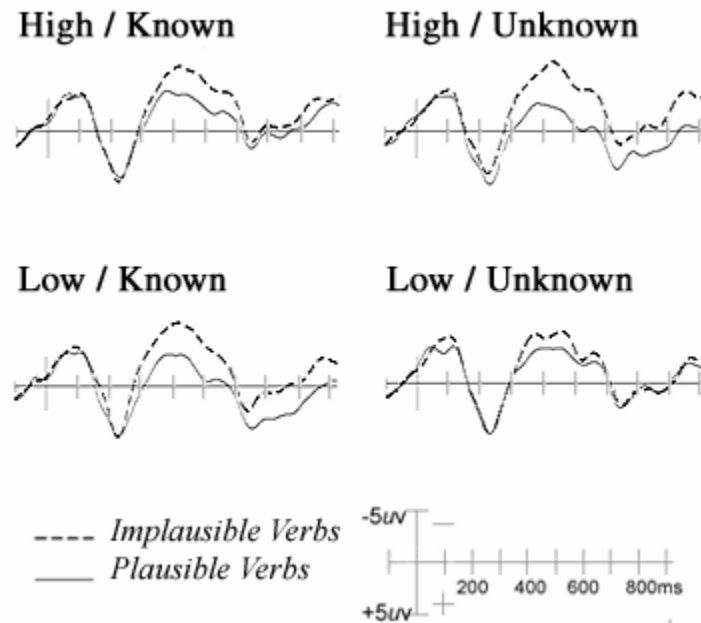


Figure 3-2. Plausibility effect for Verbs preceding Known and Unknown targets. Event-related potentials at right-medial central electrode (RMCE) are plotted for plausible and implausible verbs in each condition.

In sum, the results of the study are clear: There is a fast word learning process in adults, and it is sensitive to contextual constraint. One exposure is sufficient to rapidly integrate enough of the meaning of unknown word to know whether or not it can serve as the object of certain verbs, but only if the learning context is highly constraining. The ERP data reveal that this knowledge about a newly experienced word is relatively sophisticated and used in real-time sentence processing. Whereas previous work explored neural signatures of word learning over multiple exposures (Mestres-Misse, Rodriguez-Fornells, & Munte, 2006) or extended training (McLaughlin, Osterhout, & Kim, 2004; Ojima, Nakata, & Kakigi, 2005; Perfetti, Wlotko, & Hart, 2005; Stein et al., 2006), to our knowledge, this is the first report of the neural (electrical) correlates of one-shot word learning in adults.

Perhaps not surprisingly, we did not see modulation of the N400 component to unknown words by contextual constraint upon their initial presentation. In contrast, known words, as expected, yielded significantly smaller N400 amplitudes and larger subsequent late positivities when they completed higher constraint sentence fragments. However, the influence of initial context on lexical acquisition from the single preceding exposure of novel words were clearly evident in the modulation of the N400 amplitude to the main verb in the subsequent probe sentence. Under the strong constraint learning condition, N400 amplitudes to the probe verbs were significantly smaller if they could plausibly take the novel words as objects than if they could not. If the unknown words had initially appeared in a weakly constraining context, then there was no reliable plausibility effect observed in N400 amplitude (or any other ERP component) elicited by the probe verb. In contrast, the pattern of N400 plausibility effects for known words are present regardless of the contextual constraint of the original sentence in which they appeared, presumably because their meanings were known prior to its exposure in the current experiment.

Although a plausibility judgment was neither required nor given until after the presentation of the final word in probe sentence, the N400 amplitude modulation at the verb shows that plausibility was computed as soon as it could be, consistent with incremental views of sentence processing (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003). The presence of the plausibility effect at the verb further demonstrates that novel word knowledge was immediately coordinated with verb selectional information. In this case, participants rapidly processed information present at the verb that includes sophisticated information about appropriate semantic and syntactic

information that could be expected in the upcoming object, but only in cases where the expected meaning was highly constrained by a previous example. Thus, after a single exposure in a highly constraining context, the mental representation of an unknown word is rapidly incorporated into the lexical and grammatical systems of participants' language processing system.

In conclusion, we explore an important aspect of word learning – that of learning how to use a word. This type of information can only be received via experience, and we find that the type of experience matters. This effect can shed light on common problems encountered by all language learners, such as why L1 and L2 learners learn some words but not others, or why native speakers can encounter problems when using a thesaurus to replace common words with their apparent synonyms. In addition, this study suggests a new method in which it is possible to assess knowledge about very recently learned linguistic information using a procedure that does not require a physical response. This has many potential applications in patient populations as well as in educational contexts.

Chapter 3, in part, has been submitted as it may appear for publication; Borovsky, Arielle, Elman, Jeff, and Kutas, Marta. (2008). *Learning to use words: Event related brain potentials index single shot word learning from context*. Under review in *Cognition*. The dissertation author was the primary investigator and author of this paper.

CHAPTER 4

ONCE IS ENOUGH: N400 INDEXES SEMANTIC INTEGRATION OF NOVEL WORD MEANINGS FROM A SINGLE EXPOSURE IN CONTEXT

4.1 Abstract

The neural and cognitive correlates of word learning have become a topic of recent interest to researchers, with recent findings suggesting that the brain is able to acquire word meanings with relatively minimal training. In this study, we investigate how contextual constraint influences the integration of novel word meanings into semantic memory. We measured electrical brain activity as young adults read sentences containing known and unknown words in strongly or weakly constraining contexts. Subsequently, word knowledge was assessed by asking the participants to perform a lexical decision task, in which the known and unknown words that had initially appeared in the sentences were seen again as prime words preceding related, unrelated, and identical or synonym word targets. As expected, N400 amplitudes to target words that were preceded by (known) word primes varied as a function of prime relatedness, with more highly related words eliciting smaller N400s. Interestingly, N400 amplitudes to targets preceded by novel words that had initially appeared in highly constraining sentences also elicited this pattern of N400 modulation in relation to semantic relatedness, whereas those preceded by novel words that had initially

appeared in weakly constraining sentences did not. These results demonstrate that electrical brain activity due to one-shot contextual word learning is modulated by contextual constraint and reveals a rapid neural process that operates to integrate information about novel word meanings into the mental lexicon.

4.2 Introduction

Although improvements in word learning are most dramatic during early childhood, adult vocabularies continue to expand throughout the lifespan. It has been estimated that adults come to know between 40000 – 150000 words, whereas pre-literate five or six year old children will know only 2500-13000 words (Aitchinson, 1994; Beck & McKeown, 1991; Bloom, 2000; Pinker, 1994). Indeed, word learning continues through life and the large majority of vocabulary is acquired after childhood, unlike with other language acquisition abilities that diminish with age, such as mastery of syntax (Hakuta, Bialystok, & Wiley, 2003; Johnson & Newport, 1989; Lenneberg, 1967; Mayberry & Lock, 2003; Mayberry, Lock, & Kazmi, 2002).

Research on the factors that influence lexical acquisition in children has tended to focus on the degree to which cognitive mechanisms involved in explicit object and action name learning paradigms are specific to language learning (Bloom, 2000; Childers & Tomasello, 2002; Deak, 2000; Markman, 1992; Waxman & Booth, 2000). Adult lexical acquisition has mostly been investigated in studies of second language lexical acquisition. This research has emphasized the cognitive and neural similarities (and differences) between word learning in first and second languages (Costa & Santesteban, 2004; Francis, 1999; Halsband, 2006; Hernandez, Dapretto, Mazziotta, & Bookheimer, 2001; Illes et al., 1999;

Marian, Spivey, & Hirsch, 2003; Singleton, 1999; Tan et al., 2003). Although both areas of study have yielded important insights into how young children and bilingual adults learn words, the mode by which adults learn words in their native language is likely to differ from that of young children and adult bilinguals. For instance, studies of pre-literate children often explore word learning in explicit training contexts, such as in learning to name a novel object. In contrast, older children and adults acquire words almost entirely via incidental learning in various language contexts, especially during reading (Jenkins, Stein, & Wysocki, 1984; Nagy, Anderson, & Herman, 1987; Nagy, Herman, & Anderson, 1985; Sternberg, 1987). Given that adult word learning is so common, but differs from that of child lexical learning, it is surprising that there has been relatively little systematic work on this topic.

One major question that is of particular importance is how language context may influence lexical acquisition. It is known that under the right conditions word learning can be remarkably fast. In certain cases, only a single exposure to a novel word is sufficient for a learner – child or adult – to infer its probable meaning (Carey & Bartlett, 1978; Dollaghan, 1985; Heibeck & Markman, 1987) or understand its appropriate usage (Borovsky, Elman & Kutas, 2007). However, little is known regarding what information is rapidly acquired about novel word meanings from just a single exposure, how or the extent to which context influences this representation, and what the neural correlates of this rapid learning may be. Therefore, the main goal of our research is to explore these issues by measuring modulation of event-related brain potentials (ERPs) and reaction times due to single trial word learning from sentence contexts.

A growing body of research on the neural mechanisms of word learning has revealed that adult-like patterns of brain activity during lexical processing emerge between 13 to 20

months of age (Friedrich & Friederici, 2004; Friedrich & Friederici, 2005a, 2006; Mills, Coffey-Corina, & Neville, 1997; Torkildsen et al., 2006; Torkildsen, Syversen, Simonsen, Moen, & Lindgren, 2007). Studies of adult first and second language learners also provide evidence for rapid neural changes in young adults due to word learning in both L2 and L1 (Borovsky, Elman, & Kutas, 2007; McLaughlin, Osterhout, & Kim, 2004; Mestres-Misse, Rodriguez-Fornells, & Munte, 2006; Ojima, Nakata, & Kakigi, 2005; Perfetti, Wlotko, & Hart, 2005; Stein et al., 2006). For example, McLaughlin and colleagues (2004) compared brain responses in native French speakers to undergraduates learning French as a second language. They found that brain responses during a semantic priming task using French words are indistinguishable between native speakers and college-aged learners after only a few months of instruction. Their findings demonstrate not only that the brain may process word meanings acquired in childhood and adulthood similarly, but that modulation in neural activation to words over extended learning reflects lexical acquisition.

In addition, L1 word learning studies have suggested that even faster neural changes due to word learning are possible (Perfetti, Wlotko & Hart, 2005; Mestres-Missé, Rodriguez-Fornells & Münte, 2007; Borovsky, Elman & Kutas, 2007). For example, Mestres-Missé and colleagues (2007) found that three presentations of a novel word in progressively constraining sentence contexts can result in significant modulation of neural responses to words in context. However, in this study, since novel words were always trained in sentences presented in the same order from weakly to highly constraining, it is impossible to tell if the changes observed by the third presentation are due to the constraint of the sentence itself, or to changes that occur incrementally over several presentations. Following up on this study, Borovsky Elman and Kutas (2007) examine how contextual constraint can influence

understanding of novel word usage after only a single presentation. In this study, novel words were presented in a single highly or weakly constraining sentence context. Subsequently participants were asked to differentiate between appropriate and inappropriate usages of these novel words as objects of particular verbs.

While understanding how a word is used is an important aspect of vocabulary acquisition, it is also vital that the meaning of the word be assimilated into one's existing mental lexicon according to its relationship to other known words. For example, part of our understanding of the words "CAT", "DOG" and "CHAIR", is that "CAT" and "DOG" have many similarities and features in common that are not shared by "CHAIR". Recent work with adults has revealed that adults can gain significant knowledge of this kind of relationship between word meanings with a few exposures in sentence contexts. In previous work we have shown that only a single exposure to a novel word in a highly constraining sentence context is sufficient to impart significant understanding of word usage; this was probed by testing comprehenders' expectation of how the novel word might be used in a sentence. In this work, we ask whether such one trial learning is sufficient to enable learners to incorporate the novel word into the semantic network that connects words with related meanings. In addition, we explore if, and how, sentence context can influence this acquisition. More specifically, we use an event-related brain potential (ERP) component – the N400 – to index knowledge of word meaning via semantic priming when unknown words are initially presented in sentences that either strongly or weakly constrain their meaning.

The N400 is an ERP component that is a particularly sensitive measure of word learning. It is a negative going wave with a centroparietal maximum that peaks approximately 400ms after the onset of any potentially meaningful stimulus (Kutas &

Federmeier, 2000; Kutas & Hillyard, 1980). The N400 amplitude has been found to decrease when a word is more expected or when features associated with its meaning are more easily integrated within its surrounding context (Kutas & Federmeier, 2000; Kutas & Hillyard, 1980; Kutas & Van Petten, 1994). For example, Kutas and Hillyard (1980) recorded brainwave responses to sentence completions that were either congruent or incongruent with the context of the preceding sentence. In a sentence like: "I drink my coffee with cream and *sugar*" where the sentence ending was congruent given the sentence context, the elicited N400 response was much smaller than to sentences like "I drank my coffee with cream and *dog*" where the sentence completion was incongruent with the prior sentence context. One of the best predictors of N400 amplitude to words in sentential context is the eliciting word's cloze probability (Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984). Cloze probability is measured by determining the probability that a particular word is given in a context on a sentence completion task. For words with low cloze probabilities, the N400 is large, and the N400 decreases accordingly as cloze probability increases. This suggests that the N400 amplitude is related to a word's degree of expectancy or ease with which its meaning may be integrated with its context. Additionally, the N400 for orthographically legal and pronounceable nonwords (pseudowords) and is as large and sometimes larger than that for real words, and is larger for low frequency compared to high frequency words. At the same time the N400 is not present for true nonwords that do not have orthographically legal spellings, or are unpronounceable (Bentin, 1987; Bentin, 1985). N400 amplitude, thus, is associated with a word's meaningfulness in a given context, ranging from very small in amplitude when a word is very easily integrated or understood, to very large when the meaning of a word is unknown, or is not a word.

These findings suggest that the N400 is likely to vary with the degree to which the meaning of a newly encountered word is known. As described above, N400 amplitude in second language learners is reduced commensurate with their experience with the second language (McLaughlin, Osterhout & Kim, 2004). Mestres-Missé and colleagues (2007) also reported modulation in N400 amplitude as new words are incrementally learned across several sentences, although not when the sentential contexts did not make sense of the word. It is therefore likely that contextual constraint may also result in N400 changes that are associated with word learning. Previous work exploring how context influences rapid one-shot lexical acquisition has shown that contextual constraint does affect knowledge of word usage (Borovsky, Elman & Kutas, 2007). However, while understanding how to use a word is an important factor in language acquisition, it is arguably as important to integrate new words into an existing mental repertoire of word meanings and to understand their meanings in relation to that of other words. Here, we manipulate the degree to which a single, precise meaning of a novel word can be predicted from context by varying the cloze probability of the context in which novel words appear. Following the initial presentation of the unknown words in context we gauge successful word meaning acquisition by means of a semantic priming task.

An extensive body of research demonstrates that target words preceded, or primed, by an identical or related word (for example doctor – NURSE, or doctor - DOCTOR) are associated with both faster response times (see Neely, 1991 for a review), and reduced N400 amplitudes (Anderson & Holcomb, 1995; Bentin, McCarthy, & Wood, 1985; Brown & Hagoort, 1993; Deacon, Hewitt, Yang, & Nagata, 2000; Nobre & McCarthy, 1994; Ruz, Madrid, Lupianez, & Tudela, 2003), compared to target words preceded by words that are

unrelated in meaning, or compared to nonwords (i.e. doctor – CHAIR, or doctor – FOOP). Such effects have been generally interpreted as reflecting the semantic organization of known words in the brain (Collins & Loftus, 1975; Hutchison, 2003; Lucas, 2000; McRae, deSa, & Seidenberg, 1997; Plaut & Booth, 2000). In this study, we examine whether newly encountered words also can serve as an effective prime after only a single exposure within a sentence. In this case, the N400 modulation to a target word by a newly learned prime word is taken as an index of semantic integration of the novel word's meaning into memory. We use N400 amplitudes (to target words) to gauge how contextual constraint influences acquisition of word meaning by contrasting how these same novel words prime target words that are identical, related, or unrelated in meaning. (Relatedness is determined by the novel word's meaning, as implied in sentence context.) We can also explore how context impacts the integration of novel word meaning into semantic memory by assessing the interaction between the priming effect and contextual constraint.

4.3 Methods

4.3.1 Participants

24 college students (13 F, 11 M) were given credit or paid \$7/hr for their participation. Ages ranged between 18-30 (mean: 19.50). All participants were right-handed, native English speakers, and had no significant exposure to another language at least before the age of 12. Participants reported no history of mental illness, learning disability, language impairment, drug abuse, or neurological trauma. All participants had normal hearing and normal (or corrected to normal) vision. An additional 13 participated but were not analyzed: 5 had excessive blinking or motion artifact, 1 because of equipment

failure or experimenter error, and 5 reported a characteristic which disqualified them from analysis (4 had significant second language exposure as a child, 1 had non-normal vision.)

4.3.2 Materials

Stimuli consisted of 132 sentence pairs selected from Federmeier and Kutas (1999), and 528 word pairs selected to correspond with 132 sentence final words. Both are described in detail below:

4.3.3 Sentences

64 high constraint and 64 low constraint sentence pairs were selected from Federmeier & Kutas (1999). These pairs had previously been extensively normed to ensure adequate levels of cloze probability for high and low constraint sentences. Sentence pairs consisted of an initial sentence that set up an expectation of a meaning and item category, and a second sentence that was matched with sentence final words that were either plausible and expected known word sentence completions (Federmeier & Kutas, 1999), or unknown pseudowords, yielding 32 sentences in each of four main conditions: 1) High constraint / Known word ending, 2) High Constraint / Unknown word ending 3) Low constraint / Known word ending and 4) Low constraint / Unknown word ending. Sentences were counterbalanced such that each High and Low constraint sentence pairs appeared with both a Known and Unknown ending equally across all versions of the study, but not repeated within a subject. Known word target items consisted of words in 64 categories, and these target categories were used as the basis for selecting semantically related and unrelated prime-target pairs, described below. The sentence stimuli were counterbalanced across

versions so that all sentences appeared with both Known and Unknown word endings with equal frequency across participants. Table 4-1a includes examples of sentence stimuli.

Table 4-1. Examples of sentences and word pairs in each condition

A) Context Sentences (Context Constraint / Word Type)	
High/Known	Peter sat gaping at the centerfold.
High/Unknown	He asked his friend if he could borrow the MAGAZINE.
	Peter sat gaping at the centerfold.
Low/Known	He asked his friend if he could borrow the YERGE.
	The package was rectangular and heavy and suspiciously academic.
Low/Unknown	Bianca was disappointed that her uncle was giving her a BOOK.
	The package was rectangular and heavy and suspiciously academic.
	Bianca was disappointed that her uncle was giving her a SHUS.

B) Word Pairs (Prime – Target)			
	<u>Identical</u>	<u>Related</u>	<u>Unrelated</u>
High / Known	magazine - MAGAZINE	magazine - NOVEL	magazine- ACCIDENT
High/Unknown	Yerge - MAGAZINE	yerge - NOVEL	yerge- ACCIDENT
Low / Known	book – BOOK	book - LETTER	book - ROAD
Low/Unknown	shus – BOOK	shus - LETTER	shus - ROAD

Note: all word pairs were also paired with an equal number of pseudoword targets, not depicted in this table

4.3.4 Word-Pairs

528 word pairs were constructed for the study, consisting of a prime followed by a target word presented one stimulus at a time. Primes were the Known and Unknown words from the sentence endings described above. Three types of real word targets for each prime

were selected: 1) Synonym/Identical meaning (rabbit-RABBIT), 2) Related (rabbit-MOUSE), and 3) Unrelated (rabbit-RIBBON). Unrelated word pairs were selected to be as closely matched in number of letters and syllables, concreteness, imageability, familiarity, and frequency, as reported by the MRC psycholinguistic database (Wilson, 1988). In addition, Related word pairs were selected to be semantically and categorically related, as well as to be matched as closely as possible in concreteness, imageability, familiarity, and frequency. Highly associated word pairs were not included (like mouse-CHEESE, or bread-BUTTER), as confirmed via the Edinburgh Associative Thesaurus (Kiss, Armstrong, Milroy, & Piper, 1973). In cases involving Unknown word primes, Related, Unrelated and Synonym was determined by the implied meaning of the sentence context in which the Unknown word had previously appeared. An equal number of Nonword targets were also constructed so that “Yes” and “No” lexical decision responses were equivalent. Nonwords were constructed using the ARC Nonword database (Rastle, Harrington, & Coltheart, 2002), and were selected to be pronounceable, conform to English phonotactics, and contain between 4 and 7 letters. In any one version, participants saw each Known and Unknown prime paired with two out of three possible real word targets. Versions were counterbalanced such that each Known and Unknown prime was matched with the targets with equal frequency across versions. Table 4-1b includes examples of word pairs in the study.

4.3.5 Procedure

Participants were tested in a single experimental session conducted in a soundproof, electrically-shielded chamber and were seated in a comfortable chair in front of a monitor. Sessions consisted of two interleaved sections: sentence comprehension and priming.

In the sentence comprehension task, participants were instructed to read the sentence pairs for comprehension and to do their best to understand the sentence and words even when “nonsense” words appeared on the screen. The first sentence in each pair was presented in its entirety on the monitor, and participants were instructed to press a button to indicate that they had completed reading this sentence and were ready for the second. The second sentence was preceded by a series of crosses (500 ms duration with a stimulus-onset-asynchrony varying randomly between 300 and 800 ms) to orient the participant toward the center of the screen. Sentences were then presented one word at a time, each for 200 milliseconds with a stimulus-onset-asynchrony of 500 ms. Participants were asked to minimize blinking and movement during sentence presentation. The final target word appeared on the screen for 1400 ms.

In the priming task, participants were instructed to read every word that appeared on the screen and indicate with a button press if the target item (which always appeared in capital letters) was or was not a real word. Participants viewed two sets of prime/target pairs, and were given a 2500ms offset period to blink between pairs. Prime pair onsets were preceded by a set of fixation crosses that were randomly presented for 200-500ms. Immediately following the fixation cross, a prime word appeared for 200 ms, followed by an offset of 300ms, followed by the target word presentation for 200ms, and offset of 800ms. Participants provided a lexical decision response as soon as possible after the presentation of each target word in capital letters.

Sentence comprehension and priming tasks were interleaved as follows. Participants read 12 sentences pairs, and then completed the priming task consisting of 48 pairs, with primes being selected from the 12 sentences endings that had just been previously read.

Participants were given a break before beginning a new block of sentences. The entire experiment consisted of 11 blocks of sentence/prime sets.

At the end of the study, participants were asked to complete an old/new memory questionnaire containing 50 sentences that had appeared in the study, and 50 sentences that had not. Participants were not told at the beginning of the experiment that this questionnaire would be given, and were asked to indicate which sentences had appeared during the study, and which had not. This was given to ensure that participants sufficiently attended to the sentences during the study.

4.3.6 Electrophysiological recording

Scalp potentials were continuously recorded from 26 geodesically arranged sites using an ElectroCap with tin electrodes. Electrodes were placed at equal distances across the scalp, with positions and labels shown in Figure 4-1. A left mastoid reference was used. Potentials were digitized at a sampling rate of 250 Hz and hardware bandpass filter of 0.1-100Hz with Grass Amplifiers. Impedances were kept below 5 k Ω . The ERPs were stimulus-locked averages consisting of a 100-ms baseline and a 920 post-stimulus interval.

4.3.7 Data analysis

Data were re-referenced offline to an average left and right mastoid. Trials contaminated by eye movements, blinks, excessive muscle activity, or amplifier blocking were rejected off-line before averaging. ERPs were computed for epochs extending from 100 ms before stimulus onset to 920 ms after stimulus onset. Averages of artifact-free ERP trials were computed for the target words in the four learning conditions (High/Known,

High/Unknown, Low/Known, Low/Unknown) as well as to targets in all priming conditions (Identical/Synonym targets, Related targets, and Unrelated targets for each of the four main conditions High/Known, High/Unknown, Low/Known, Low/Unknown) after subtraction of the 100 ms pre-stimulus baseline

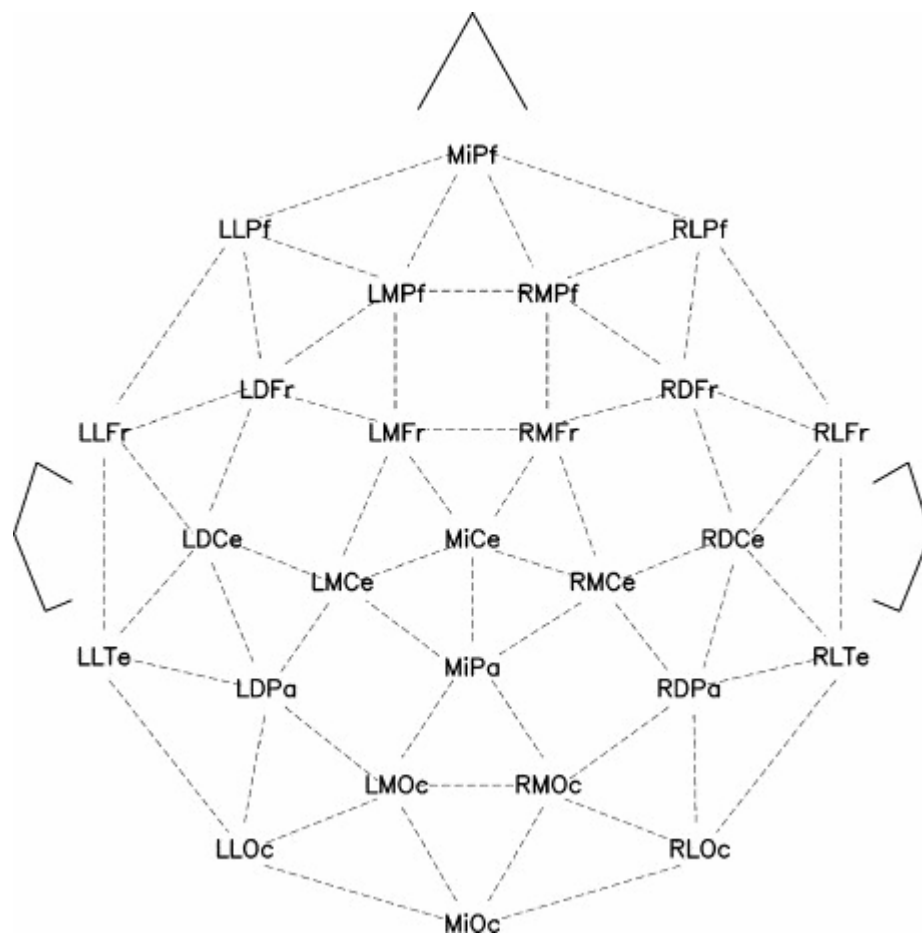


Figure 4-1. Diagram of the electrode positions and labels

4.4 Results

4.4.1 Behavioral performance

During the task, participants made lexical decisions for words that were identical, related, or unrelated in meaning to a prime word. Mean accuracy scores on this task are shown in Table 4-2. Since accuracy was near ceiling, with the lowest accuracy in any single condition being 93%, we did not statistically analyze effects of accuracy. Mean reaction times are also shown in Table 4-2 and Figure 4-2. A three factor 2x2x3 ANOVA was carried out with factors of Word type (Unknown and Known) x Constraint (High and Low) x Prime relationship (Identical, Related and Unrelated). A main effect of Prime was found [$F(2, 274)=11.88$, $p< 0.0001$], with post-hoc Tukey tests revealing that Identical targets elicited faster responses than every other condition. No difference was observed between related and unrelated conditions overall. There was no other main effects of Word Type [$F(1, 274)=1.99$, $p=0.160$] or Constraint [$F<1$]. An interaction of Prime x Type was also observed [$F(2, 274)=4.22$, $p<0.016$]. Follow-up Tukey tests revealed that this interaction was driven by targets that were preceded by Identical Known words eliciting faster responses than words in any other condition. No other significant two- or three-way interactions were observed. Although no significant three way interaction was observed, Tukey tests pairwise comparisons were conducted to examine the relationships between Identical, Related and Unrelated meanings in each of the four prime conditions: Known/High, Unknown/High, Known/Low, and Unknown/Low. These analyses revealed that targets preceded by Known/High and Known/Low word primes elicited faster reaction times when preceded by a

word identical in meaning, compared to a related or unrelated word. On the other hand, targets preceded by Unknown words did not elicit priming effects in any condition.

In order to further examine the priming relationship across all conditions, we also carried out items analyses with a three factor ($2 \times 2 \times 3$) repeated measures ANOVA with identical factors to the analysis carried out above. Like the by-subjects analysis, these tests also revealed main effects of Word Type [$F(1,773)=20.72$, $p<0.0001$], with targets preceded by Known words eliciting faster responses, and a main effect of Prime Relationship [$F(2,773)=69.88$, $p<0.0001$], with targets preceded by primes that are Identical in meaning eliciting faster reaction times. A two-way interaction of Word Type x Prime Relationship [$F(2,773)=26.15$, $p<0.0001$] was also observed, with follow-up Tukey tests indicating that this interaction was driven by targets that were preceded by Identical Known words eliciting faster responses than words in any other condition. No other significant two- or three-way interactions were observed. Although no significant three way interaction was observed, Tukey tests pairwise comparisons were conducted to examine the relationships between Identical, Related and Unrelated meanings in each of the four prime conditions: Known/High, Unknown/High, Known/Low, and Unknown/Low. These analyses revealed that targets preceded by Known/High and Known/Low word primes elicited faster reaction times when preceded by a word identical in meaning, compared to a related or unrelated word. On the other hand, targets preceded by Unknown words did not elicit priming effects in any condition.

Table 4-2. Mean reaction times (ms) and mean percentage of correct responses for priming task in all conditions

	Real Word Primes		Novel Word Primes	
	High Constraint	Low Constraint	High Constraint	Low Constraint
<i>Percent correct</i>				
Synonym / ID	99 (0.6)	99 (1.9)	97(6)	98(2.1)
Related	97 (2.4)	93 (4.1)	94(4.3)	95(3.5)
Unrelated	93 (6.8)	96 (3.2)	95(3.4)	94(3.8)
<i>Reaction time</i>				
Synonym / ID	512 (80)	488 (82)	543 (77)	553 (76)
Related	568 (87)	561 (72)	567 (79)	570 (83)
Unrelated	586 (79)	578 (75)	571 (75)	567 (79)

Note: Standard deviations are reported in parenthesis.

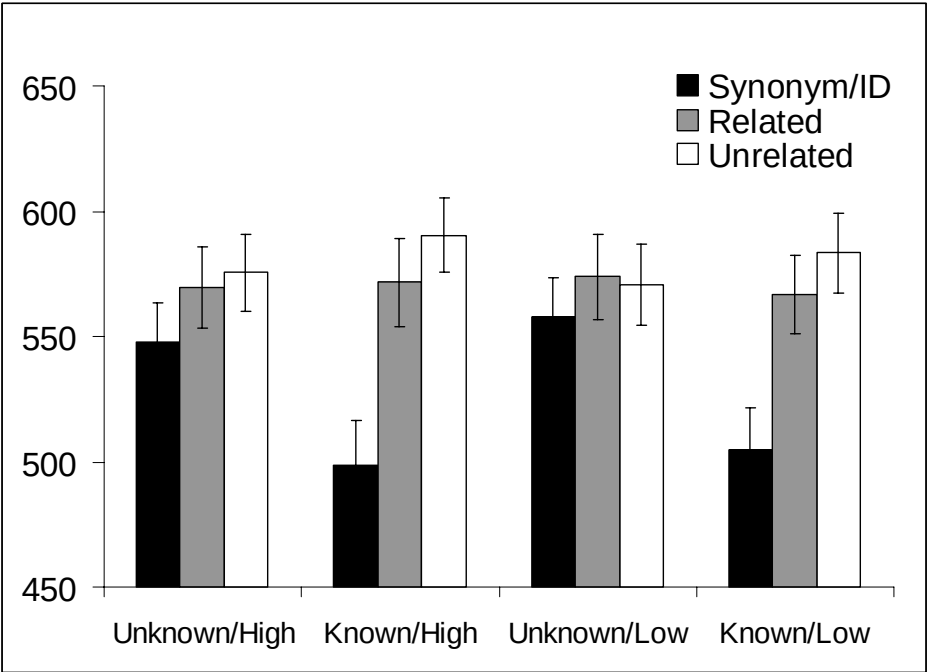


Figure 4-2. Mean reaction time to target items on priming task

4.4.2 ERP data: N400 amplitude

4.4.2.1 Context sentence endings

We analyzed artifact-free ERP responses to sentence final target words in four conditions: Known/High, Known/Low, Unknown/High and Unknown/Low. ERPs to sentence final endings are shown in Figure 4-3. N400 mean amplitude was measured between 250-500ms post final word onset at four centro-parietal electrode sites (RMCe, LMCE, MiCe, MiPa) where N400 effects are typically largest. A two-factor repeated measures ANOVA with factors of Word Type (Known and Unknown) and Constraint (High and Low) revealed no main effects of Constraint [$F(1,23)=1.47$, $p=0.2384$], and a marginal effect of Word Type [$F(1,23)=3.78$, $p=0.0642$], with novel words eliciting larger N400 responses. A Constraint x Word Type effect was significant [$F(1,23)=26.36$, $p<0.0001$], with follow-up Tukey tests revealing that this effect was driven by opposite N400 effects of constraint on Known and Unknown words: Known words that appeared in low constraint contexts yielded larger N400s (compared to other Known targets), whereas Unknown words in High constraint contexts yielded larger N400s, (compared to other Unknown targets) (all $p<0.05$).

4.4.2.2 Priming task

Grand average ERPs to target words in the four main prime word conditions (Known/High, Known/Low, Unknown/High, Unknown/Low) are shown in Figure 4-4 and Figure 4-5 at all electrode sites and in Figure 4-6 at a single medial central electrode. As can be seen from this figure, an effect of Target type is seen via modulation of the negative going peak from 250-500ms (N400) in all Prime conditions, except for Unknown/Low words. N400 mean amplitude was measured between 250-500ms post target word onset at four centro-

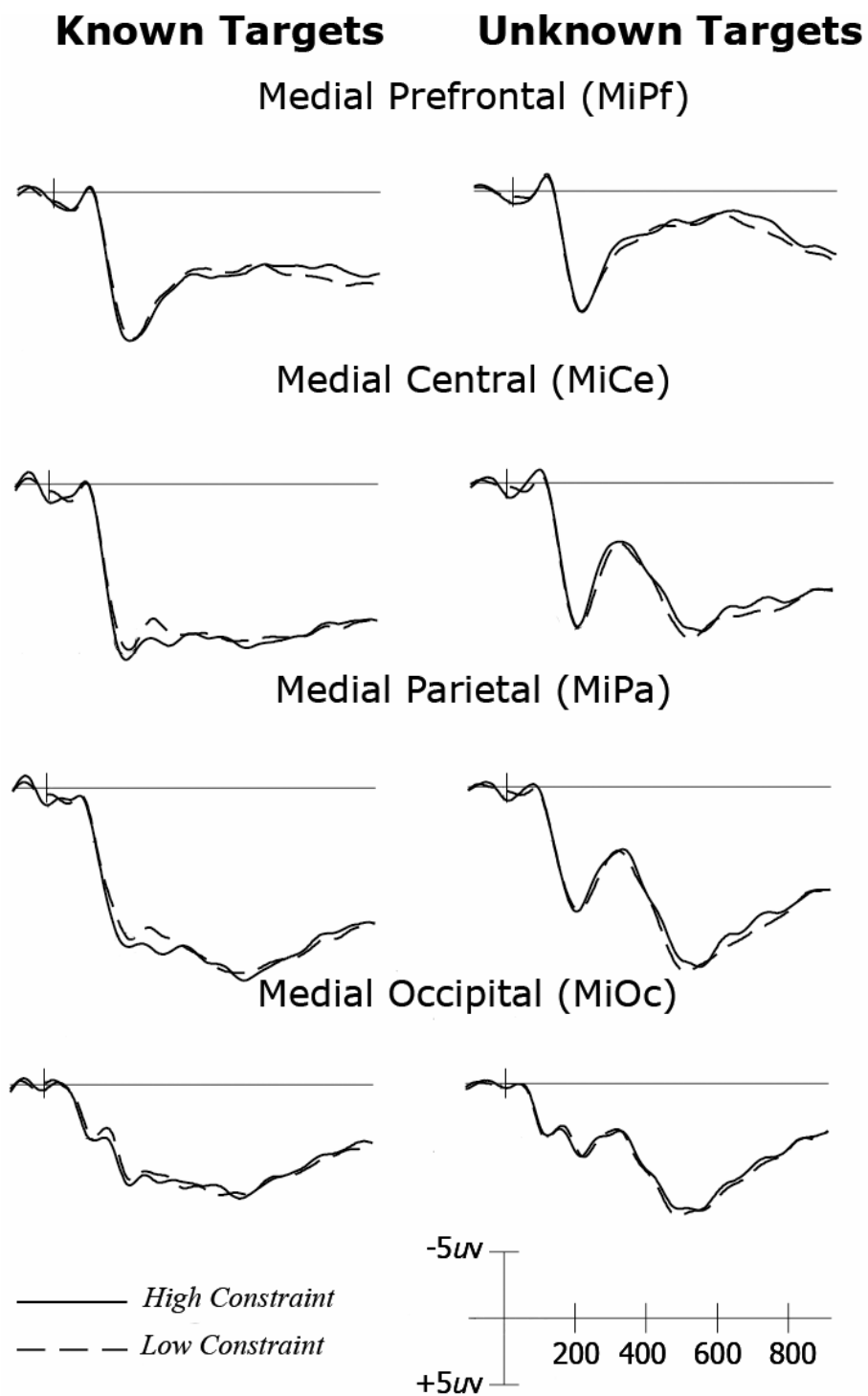


Figure 4-3. Grand average ERPs to known and unknown target words in context sentences at medial electrode sites.

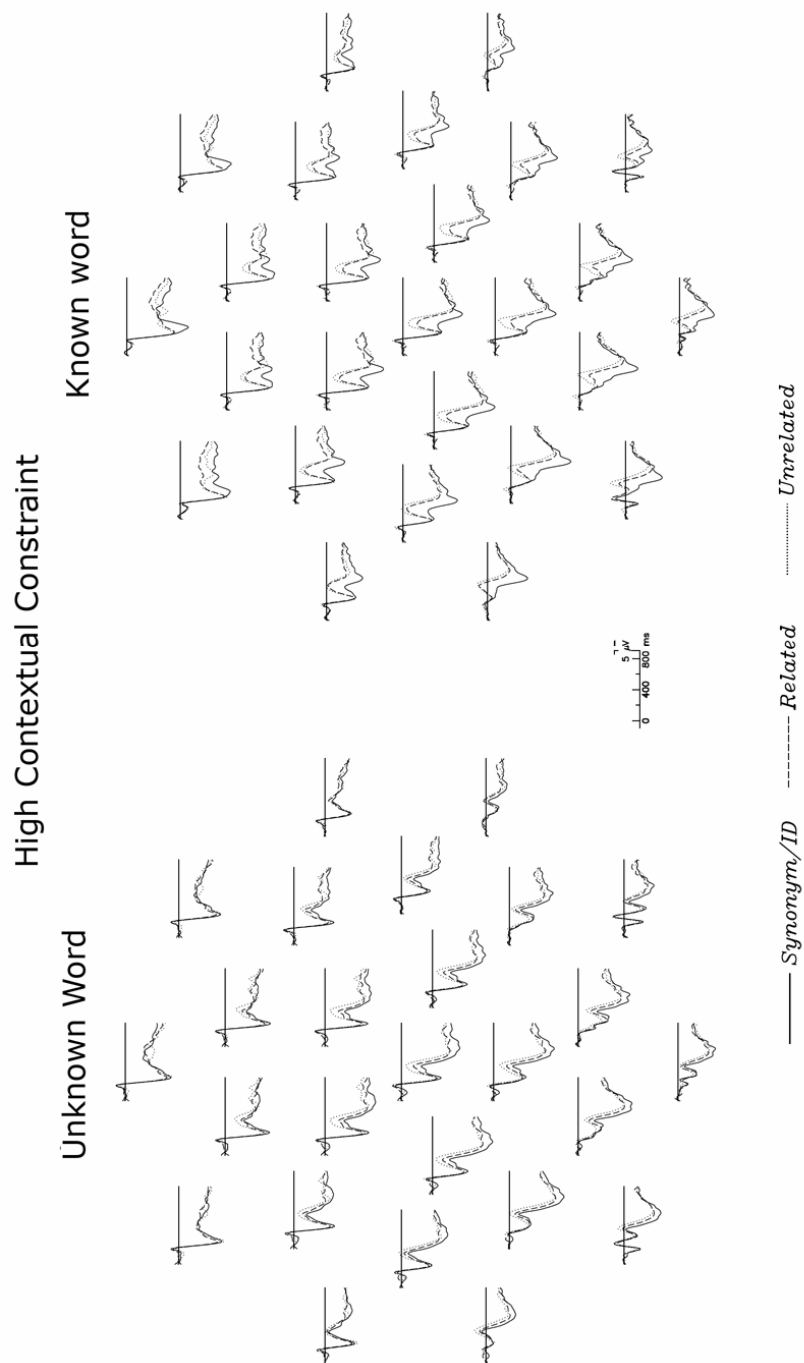


Figure 4-4. Grand average ERPs across all subjects and electrode sites for targets in the priming task that were initially preceded by Known and Unknown words that initially appeared in highly constraining sentence contexts.

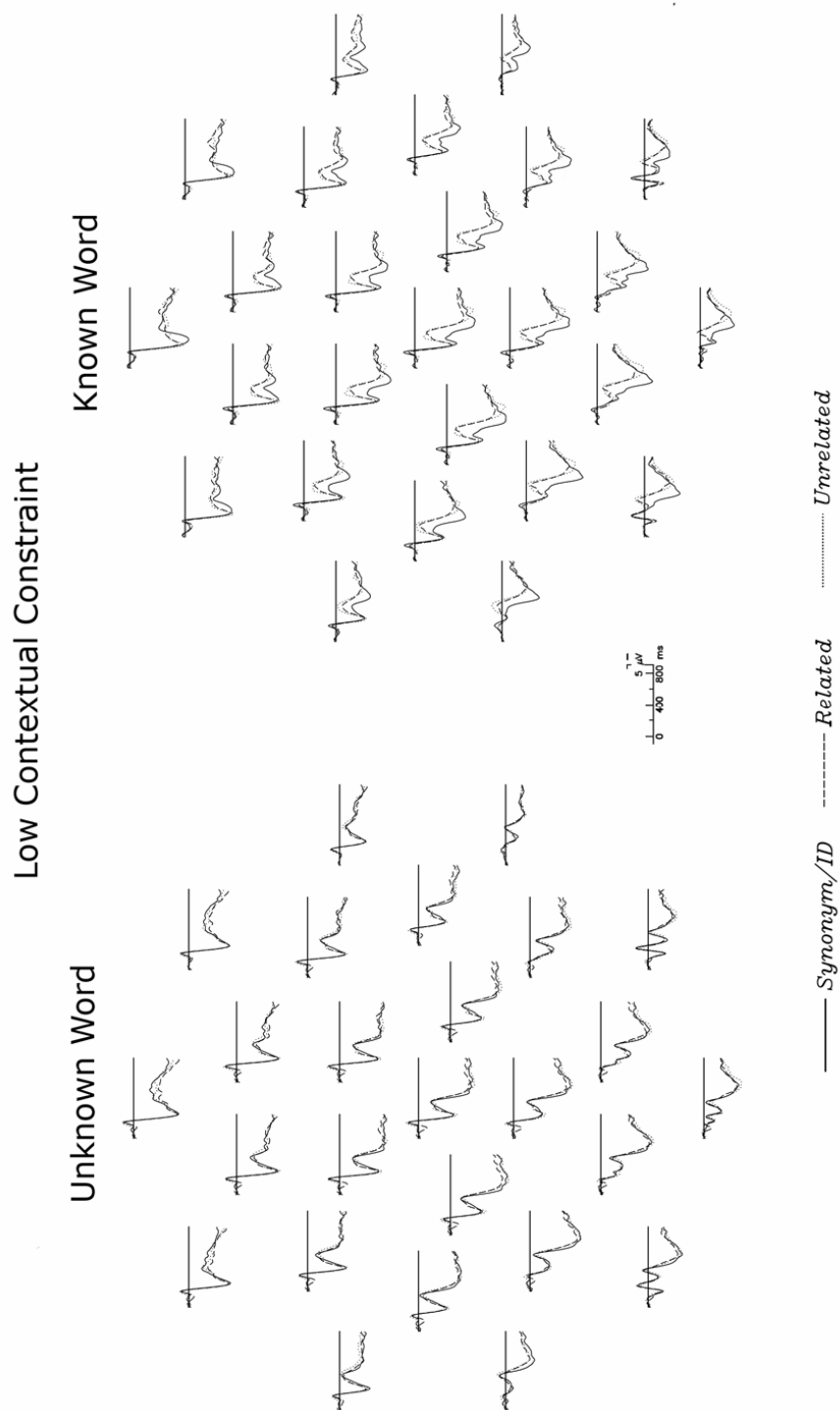


Figure 4-5. Grand average ERPs across all subjects and electrode sites for targets in the priming task that were initially preceded by Known and Unknown words that initially appeared in weakly constraining sentence contexts

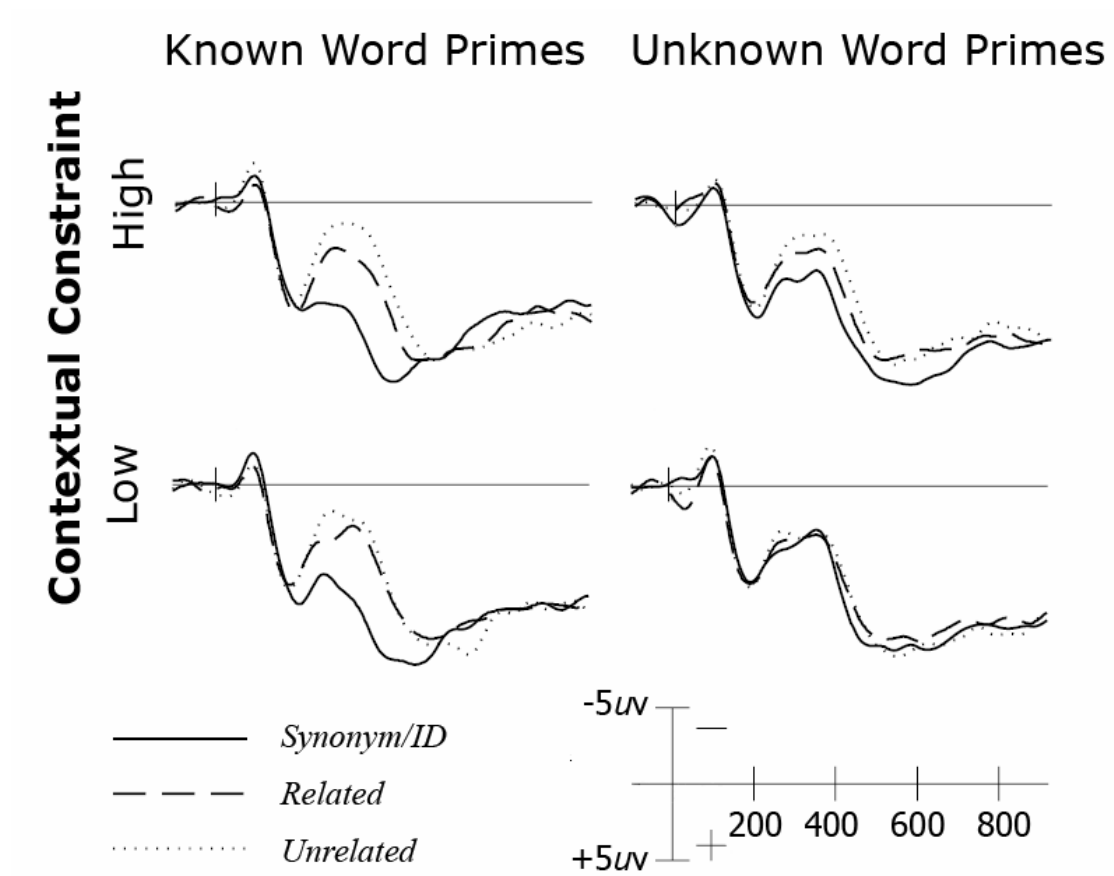


Figure 4-6. Grand average ERPs to target words in priming task at the vertex electrode (MiCE).

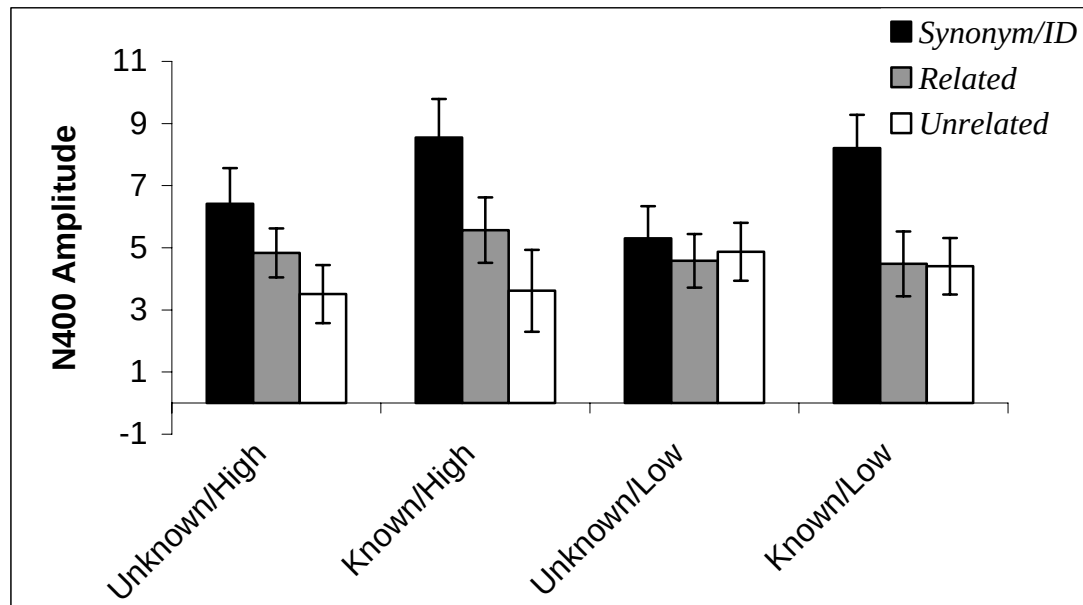


Figure 4-7. N400 Mean amplitude measured from 250-500ms averaged across four electrode sides, MiCE, RMCE, LMCE, MiPA for target words across four prime conditions

parietal electrode sites (RMCE, LMCE, MiCE, MiPA) where N400 effects are typically largest, and these values are shown in Figure 4-7. A three-factor repeated measures ANOVA was conducted with factors of Prime-Word Type (Known or Unknown), Prime-Constraint (High or Low) and Target relationship (Synonym/ID, Related, Unrelated), using Greenhouse-Geisser univariate epsilon values.

This analysis revealed a significant effect of Word Type [$F(1,23)=5.4990$, $p=0.02$], with Unknown words eliciting larger N400 amplitudes, and Target [$F(1.8922, 43.522)=32.439$, $p<0.0001$], with Synonym/ID targets eliciting smaller N400 amplitudes, but no main effect of Constraint [$F(1,23)<1$]. There was also an interaction of Constraint x Prime [$F(1,23)=6.2911$, $p=0.0196$], but no other two- or three-way interactions were significant. Preplanned pairwise

repeated measures ANOVA comparisons were then conducted to compare mean N400 amplitude between Related, Unrelated and Synonym/ID targets in each of the four main Prime word conditions. The results of these comparisons are shown in Table 4-3. As seen from this table, significant priming effects were observed in all prime conditions, except for Unknown prime words that initially appeared in Low constraint contexts.

Table 4-3. F-values from pairwise ANOVAs comparing mean amplitude N400 to related, unrelated, and synonym/ID targets in four main prime conditions

		<u>Synonym/ID</u>	<u>Related</u>	<u>Unrelated</u>
Known/High	Synonym/ID	--	14.92**	30.22***
	Related	14.92**	--	11.17**
	Unrelated	30.22***	11.17**	--
Known/Low	Synonym/ID	--	27.80***	23.69***
	Related	27.80***	--	Ns
	Unrelated	23.69***	ns	--
Unknown/High	Synonym/ID	--	6.22*	32.24***
	Related	6.22*	--	4.61*
	Unrelated	32.24***	4.61*	--
Unknown/Low	Synonym/ID	--	ns	ns
	Related	ns	--	ns
	Unrelated	ns	ns	--

* - $p < 0.05$

** - $p < 0.01$

*** - $p < 0.0001$

4.4.3 ERP Data: N400 distribution

The amplitude analyses conducted above were done with a limited set of electrodes over scalp sites where N400 effects are typically largest. One possibility is that N400 effects might be differentially distributed across the scalp for each of the experimental conditions.

Below we conduct additional analyses on sentence final words in the context sentences and target words in the priming task using a larger set of electrodes in order to examine the distribution of the observed N400 effects. In both analyses, the N400 was measured in the typical 250-500ms time window across 16 electrodes classified according to their position on the scalp. These distributional factors were: Hemisphere (Right or Left), Anteriority (Prefrontal, Frontal, Central, Occipital), Laterality (Medial or Lateral). The electrodes used in this analysis were: LLPf, RLPf, LMPf, RMPf, LLFr, RLFr, LMFr, RMFr, LLTe, RLTe, LMCE, RMCE, LLOc, RLOc, LMOc, RMOc (see Federmeier, Mai, & Kutas, 2005). Below, the analyses for the sentence final words, and then primes are reported:

4.4.3.1 Context Sentences

A repeated measures ANOVA was conducted with experimental factors of Word Type (Known and Unknown) x Constraint (High and Low), and distributional electrode factors of Hemisphere (Right and Left), Anteriority (Prefrontal, Frontal, Central, Occipital) and Laterality (Medial and Lateral). Main effects were observed for the following factors: Word Type [$F(1,23)=621.16$, $p<0.0001$], Word Type x Constraint [$F(1,23)=101.42$, $p<0.0001$], Hemisphere [$F(1,23)=9.05$, $p=0.0027$], Anteriority [$F(3,69)=2169.3$, $p<0.0001$], and Laterality [$F(1,23)=6117.6$, $p<0.0001$]. Interactions of distributional factors included: Hemisphere x Anteriority [$F(3,66)=5.31$, $p=$], and Anteriority x Laterality [$F(3,69)=629.25$, $p<0.0001$], and were driven by a tendency of the N400 to be larger over left occipital sites and larger over lateral sites at all regions except over prefrontal sites. An interaction of Word Type x Constraint was observed [$F(1,23) = 12.42$, $p=0.0004$] that was driven by no observed N400

constraint effects for novel words, while the N400 was larger for known words that appeared in weakly constraining contexts. No other interactions were found.

4.4.3.2 Priming task

A repeated measures ANOVA was conducted with experimental factors of Word Type (Known and Unknown) x Constraint (High and Low), and Prime Relatedness (Synonym/ID, Related, and Unrelated), and distributional electrode factors of Hemisphere (Right and Left), Anteriority (Prefrontal, Frontal, Central, Occipital) and Laterality (Medial and Lateral). Main effects were observed for the following factors: Word Type [$F(1,23)=50.06$, $p<0.0001$], Prime [$F(2,46)=97.83$, $p<0.0001$], Anteriority [$F(3,69)=63.58$, $p<0.0001$], and Laterality [$F(1,23)=507.19$, $p<0.0001$]. A significant interaction of distributional factors Anteriority x Laterality [$F(3,69)=55.88$, $p<0.0001$] was driven by a tendency for the N400 to be larger over lateral sites except over prefrontal electrodes. Other interactions observed were: Word Type x Prime [$F(2,46)=42.77$, $p<0.0001$], Constraint x Prime [$F(2,46)=8.393$, $p=0.0002$], Word Type x Constraint x Prime [$F(2,46)=3.472$, $p=0.0311$], Word Type x Anteriority [$F(3,69)=11.49$, $p<0.0001$], Prime x Anteriority [$F(6,138)=2.91$, $p<0.0079$], Prime x Laterality [$F(2,46)=6.75$, $p=0.0012$]. No interactions between the three experimental factors together (Word Type x Constraint x Prime) and any individual or combination of distributional factors was observed.

4.4.4 Relationships between N400 and RT

Since we found that N400 priming effects were not coincident with reaction time measure of priming in the Unknown word conditions, we decided to further explore the

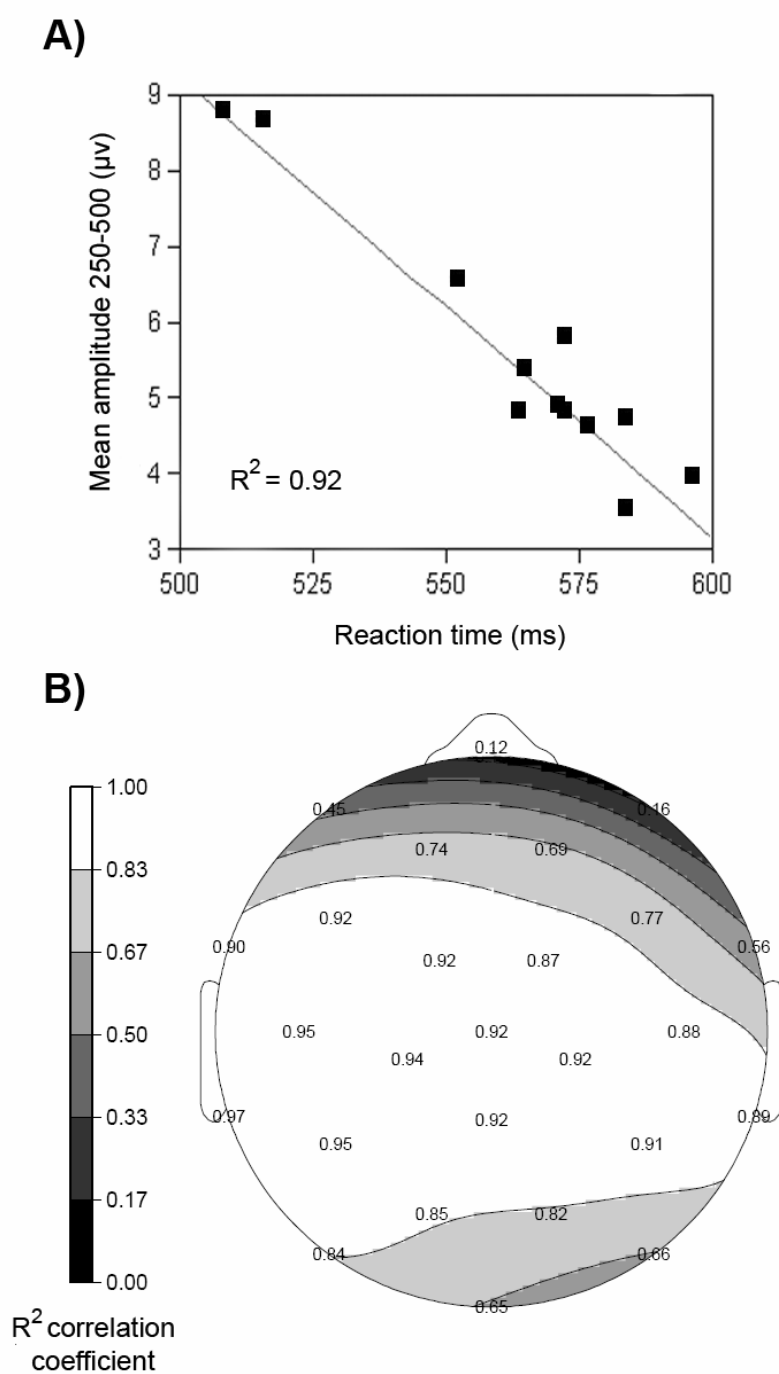


Figure 4-8. Correlations between reaction time and N400 amplitude for all priming conditions.

relationship between N400 amplitude and RT in the priming conditions. We examined the association between N400 and reaction time across all conditions by measuring the correlation between average N400 amplitude from 250-500ms at each electrode site in all 12 priming conditions (Known/High/ID, Known/High/Related, Known/High/Unrelated, Known/Low/ID, Known/Low/Related, Known/Low/Unrelated, Unknown/High/ID, Unknown/High/Related, Unknown/High/Unrelated, Unknown/Low/ID, Unknown/Low/Related, Unknown/Low/Unrelated) with that of reaction time for each priming condition. Figure 4-8a plots the mean amplitude for N400 across average reaction time at a single vertex electrode, while Figure 4-8b displays the R^2 correlation coefficients for this relationship across all electrodes on the scalp. As can be seen from these figures, N400 amplitude and reaction time were strongly negatively associated over a broad region on the scalp, such that faster reaction times were associated with more positive N400 amplitudes.

4.5 Discussion

This study explored the neural correlates of the rapid acquisition of unknown word meanings in young adults' native language. Our goal was to understand how sentential constraint influences the integration of novel word meanings into the existing "mental lexicon" after just a single exposure. This question was motivated in part by previous findings that adults can rapidly acquire knowledge of lexico-semantic relationships between words after several presentations in context, and that a single presentation of a novel word can be enough to impart significant information about its appropriate usage. In this study, we used a priming task to ask if the information that is rapidly integrated about novel word meanings includes not only information about how that word can and cannot be used but

also its lexico-semantic relationships with other known words. We examined both behavioral and ERP responses as adult participants performed a lexical decision task where known and unknown words that had previously appeared in strongly or weakly sentence contexts served as primes for known words that were either related, unrelated, or were synonyms or identical in meaning to the prime word.

Interestingly, the behavioral results did not reveal evidence of priming between novel word meanings and their related or synonymous targets. This result alone would suggest that no word learning occurred, regardless of how informative the prior sentential context might have been. However, the electrophysiological data indicate a very different outcome.

Known word primes produced significant N400 priming effects that replicate a well established result: N400 amplitudes for targets are smaller when those words are preceded by words that are identical or related in meaning, relative to unrelated primes. Although Unknown word primes produced no facilitation in the reaction time to identical or related targets, these primes did lead to a reduction in the N400 amplitude of targets that were semantically related to them. Obviously, semantic relatedness between an Unknown prime and real word target could only have been inferred from the sentence context in which the Unknown word previously appeared. Notably, however, this effect only occurred for Unknown words that had initially appeared in a strongly constraining context.

One possibility is that the priming effect in novel words may have arisen from different cognitive and/or neural mechanisms than those responsible for priming effects seen for known words. However, we did not find evidence that the distribution of the N400 for priming effects for known and unknown words was different. In addition, there was a

high correlation between reaction time and N400 amplitude across all experimental conditions. Both of these results do not support the idea that different neuron/cognitive mechanisms were responsible for the priming effects in the novel word conditions. However, ERPs are not a particularly sensitive measure of spatial resolution, and different neural generators can be responsible for waveforms that are identical in distribution, so it is still possible that spatial differences do exist between known and unknown word priming and that our method was not sensitive enough to measure them. Future work is necessary to explore this issue.

Previous work has suggested that adults can integrate and organize information about word meanings after a number of weeks of second language instruction (McLaughlin, Osterhout, & Kim, 2004; Stein et al., 2006), and even more rapidly in adult's first language, such as after only an hour of study of word definitions (Perfetti, Wlotko, & Hart, 2005) or after three presentations in sentential context (Mestres-Misse, Rodriguez-Fornells, & Munte, 2006). The present results extend these findings by showing that in some cases a single exposure of a novel word in a strongly constraining sentence context is sufficient to convey significant information about its meaning, and that there is a very fast neural process which enables the integration of this information into semantic memory.

Basically, while we found that the N400 was sensitive to acquisition of novel word meaning, we did not find accompanying behavioral evidence of priming, despite there being a high correlation between reaction time and N400 amplitude across all experimental conditions. Previous work by McLaughlin and colleagues (2004) also found a similar dissociation: no improvement in sensitivity of distinguishing real from non-words in a lexical decision task, despite observing significant changes in N400 with additional experience in

learning a second language. This suggests that the ERP method may be particularly sensitive to acquisition of word meaning information that standard lexical priming reaction time measures apparently do not reveal. This may also provide an explanation for recent findings that find that lexicalization of novel words can be slow, or occur only after sleep (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). This work had used reaction time measures to explore novel word meaning acquisition, and, as we observed – these lexical priming measures appear relatively insensitive to acquisition of novel word meanings. It appears that an ERP measure of word learning may reveal evidence of novel word integration into semantic memory even in the absence of a nap.

In summary, this is the first study to reveal that adults can rapidly integrate novel word meanings into their mental dictionaries after only a single presentation. Our results add to previous findings that the N400 component is sensitive to meaning acquisition of word meanings after multiple exposures, and understanding of how a word can and cannot be used grammatically after only a single exposure in a highly constraining context. With this study, we add to a growing body of evidence that the rapidly acquired information about novel words is highly detailed and sophisticated – and includes information about both its meaning and usage in sentence contexts.

Chapters 4, appears in preparation for publication by Borovsky, Arielle; Elman, Jeff and Kutas, Marta. The dissertation author was the primary investigator and author of this paper.

CHAPTER 5

GETTING IT RIGHT: WORD LEARNING ACROSS THE HEMISPHERES

5.1 Abstract

The brain is apparently able to acquire significant information about an unknown word's meaning from a highly constraining sentence context with minimal training. In this study, we investigate whether the nature of this novel information differs between the cerebral hemispheres. Young adults first read sentences containing previously known and unknown words in strongly or weakly constraining contexts. Subsequently, word knowledge was assessed by asking participants to perform a lexical decision task in which known and unknown words that had initially appeared in the sentence contexts served as visual prime words to unrelated, identical or synonym word targets that appeared randomly in the right or left visual field. As expected, N400 amplitudes to target words that were preceded by known word primes varied as a function of prime-target relatedness. Identical targets yielding smaller N400 amplitudes regardless of visual field of presentation. N400 amplitudes to target words that were preceded by novel word primes that had initially appeared in highly constraining sentences also exhibited this same pattern of modulation, but only with left visual field/right hemisphere presentation. N400 priming effects were not observed for targets preceded by novel word primes that had been presented in weakly constraining sentence contexts in either visual field. These results indicate that rapid neural changes in

electrical activity due to single-shot contextual word learning are driven primarily by the right hemisphere.

5.2 Introduction

How do adult brains integrate new word meanings into a well-developed mental lexicon? Recent evidence indicates that electrical brain activity in young adults can reflect the acquisition of novel word meaning in their first and second languages. In particular, the N400 of the event related brain potential (ERP) has been shown to be particularly sensitive to acquisition of various dimensions of word meaning, including knowledge of a novel word's appropriate usage as the object of a verb, and its semantic/associative relationship to other words. More specifically, lexical acquisition in native (L1) and second language (L2) can result in rapid changes in electrical brain activity over the course of several weeks of L2 instruction (McLaughlin, Osterhout, & Kim, 2004), over several trials, or over the course of a few minutes in L1 (Mestres-Misse, Rodriguez-Fornells, & Munte, 2006; Perfetti, Wlotko, & Hart, 2005), and even following a single exposure (Borovsky, Elman, & Kutas, 2007). Taken together, these findings implicate a neural mechanism that allows the brain to acquire the meaning of a new unknown word rapidly, and indicate that the N400 brainwave component is sensitive to neural effects of this process.

Despite the growing number of studies that document the neural *effects* of word acquisition, less is known regarding the neural *mechanisms* involved in the representation and acquisition of novel word meanings. One possibility is that the cerebral hemispheres may play different roles in word learning. In fact, as we review below, there is considerable evidence that the two hemispheres of the brain have differential capacities to encode and

represent verbal information. On the other hand, relatively little is known about the relative role of the different hemispheres in word meaning acquisition. One way this question can be explored is via the lateralized presentation of stimuli. This method involves presenting stimuli to either the right or left visual field (RVF or LVF) while a participant fixates a central point. The lateralized presentation takes advantage of a unique aspect of the wiring of the visual system, namely, that the majority of information from each visual field is initially sent to the contralateral hemisphere, before crossing to the other hemisphere via the various cerebral commissures. This small difference in the availability of lateralized information to the two cerebral hemispheres offers a window into the differential capabilities and involvement of the two hemispheres in processing.

One way in which the effects of word learning can be studied is via electrophysiological techniques that measure time-locked brainwave responses to stimuli of interest, or event-related potentials (ERPs). Recently, scientists have found a specific brainwave component, the N400, to be useful for this purpose. The N400 is a negative going wave with a centroparietal maximum that peaks approximately 400ms after the onset of any potentially meaningful stimulus (Kutas & Federmeier, 2001; Kutas & Hillyard, 1980). The N400 amplitude has been found to decrease when a word is more expected or when features associated with its meaning are more easily integrated within its surrounding context (Kutas & Hillyard, 1980). Additionally, the N400 for pronounceable nonwords is sometimes larger than that of real words (depending on their frequency of usage), but is not present for true nonwords that do not have orthographically legal spellings, or are unpronounceable (Bentin, 1987; Bentin, 1985). Words that are preceded by other words that are meaningfully related, or identical also elicit smaller N400 amplitudes (Anderson &

Holcomb, 1995; Bentin, McCarthy, & Wood, 1985; Brown & Hagoort, 1993; Deacon, Hewitt, Yang, & Nagata, 2000; Nobre & McCarthy, 1994; Ruz, Madrid, Lupianez, & Tudela, 2003). Therefore, N400 amplitude is associated with a word's meaningfulness in a given context, ranging from very small in amplitude when a word is very easily integrated or understood, to very large when the meaning of a word is unknown.

A growing body of evidence suggests that the N400 is also sensitive to acquisition of word meaning, such that its amplitude is reduced when a novel word meaning that has been successfully acquired is used appropriately, or is presented in conjunction with a known word synonym (Borovsky, Elman, & Kutas, 2007; Mestres-Misse, Rodriguez-Fornells, & Munte, 2006; Perfetti, Wlotko, & Hart, 2005).

Other ERP components also have been associated with a number of language-related and perceptual operations that could be sensitive to novel word processing. The visual N1 component is an early negative component that is largest over posterior electrode sites, and has been linked to early visual attentional processes. Its amplitude has been linked to attention and awareness to visual stimuli, with larger amplitudes indicating greater detection success or visuospatial attentional focus (Hillyard & Anllo-Vento, 1998). The late positive complex (LPC) is a late positivity occurring 500-900 ms after stimulus onset at centro-parietal sites that is thought to reflect semantic retrieval processes (Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991), especially those involved with conscious integration of meaning (Duzel, Yonelinas, Mangun, Heinze, & Tulving, 1997).

One paradigm in which ERP studies of word meaning acquisition have measured lexical knowledge is via semantic priming tasks in which novel words appear as primes to target words that vary in semantic relatedness to the (inferred) novel word meaning. These

studies take advantage of the well-studied, semantic priming effect (Meyer, 1970). This effect occurs when target words that have been preceded, or primed, by an identical or related word (for example doctor – NURSE, or doctor - DOCTOR) yield faster response times to the target word (for review, see Neely, 1991), and/or reduced N400 amplitudes (Anderson & Holcomb, 1995; Bentin, McCarthy, & Wood, 1985; Brown & Hagoort, 1993; Deacon, Hewitt, Yang, & Nagata, 2000; Nobre & McCarthy, 1994; Ruz, Madrid, Lupianez, & Tudela, 2003), compared to words that are unrelated in meaning, or are nonwords, (i.e. doctor – CHAIR, or doctor – FOOP). Such effects have been generally interpreted as reflecting the semantic organization of known words in the mental lexicon (Collins & Loftus, 1975; Hutchison, 2003; Lucas, 2000; McRae, deSa, & Seidenberg, 1997; Plaut & Booth, 2000). In addition, N400 studies of lexical acquisition have found that known target words that are semantically related or synonyms of preceding (and thus priming) novel words (whose meaning has been acquired to some extent) are characterized by smaller N400 amplitudes than are known target words that are unrelated to the novel word primes. These findings indicate that the brain seems to learn to appreciate relationships between novel word meanings and other known words in first and second language learning relatively quickly, and that this learning/knowledge is evident in time-locked, average electrophysiological measures.

How might this knowledge be represented across the hemispheres, and what role might each play during the learning process? Below we briefly review the literature on hemispheric involvement in encoding and representation of verbal knowledge.

5.2.1 The role of the cerebral hemispheres in encoding verbal information

In the adult neuropsychological literature a left hemisphere (LH) dominance for processing and comprehension of verbal material has been established via studies of individuals who have received lesions to a single side of the brain. In contrast, findings from intact adults have shown that both hemispheres are involved in encoding and representation of linguistic material. At the same time, this evidence suggests that each hemisphere is differentially involved in the nature of information represented and encoded. For instance, the RH tends to encode visual information based on physical features, whereas LH visual encoding occurs at a more conceptual and abstract level. For example, a number of priming and visual perception experiments have suggested that the RH tends to be much more sensitive (than the LH) to changes word form, e.g., case changes or letter size, such that two instances of a word or letter that vary in form (K vs k, or DOG vs dog) result in reduced, but still existent, priming effects compared to the LH (Marsolek, 1995; Marsolek, Kosslyn, & Squire, 1992; Marsolek, Nicholas, & Andresen, 2002; Marsolek, Schacter, & Nicholas, 1996; Marsolek, Squire, Kosslyn, & Lulenski, 1994). It seems then that both hemispheres can encode verbal information conceptually, since they both show evidence of conceptual priming that ignores the pure physical form properties of words, but there is a LH advantage for this type of conceptual encoding. Additionally, these studies suggest that the RH encodes form-specific information about words. These and other findings have led a number of researchers to suggest that the RH's superiority in visuospatial information processing may lead the RH to extract primarily physical information about word forms, whereas the LH's capacity for verbal processing allows for improved encoding and integration of conceptual information of word forms.

5.2.2 The role of the cerebral hemispheres in the representation of verbal information

In contrast with findings from the encoding literature, studies of semantic representation have provided wide evidence that both cerebral hemispheres represent significant, albeit different, semantic and categorical information about word meanings (for review, see Beeman & Chiarello, 1998). These studies suggest that the RH is specifically involved in the processing of more distant, or weakly related semantic information. For example, ambiguous word meanings (RIVER - BANK vs MONEY - BANK) (Burgess & Simpson, 1988), weakly associated category members (DEER – PONY) (Chiarello, Burgess, Richards, & Pollock, 1990) and weakly related associates (Beeman, Friedman, Grafman, & Perez, 1994) are all more strongly primed in the RH than LH. These hemispheric differences in semantic representation have been characterized as being “focused”, “fine” or “localized” in the LH and “coarse”, “diffuse”, or “distributed” in the RH (Beeman, Friedman, Grafman, & Perez, 1994; Chiarello, 1998; Grose-Fifer & Deacon, 2004). (Beeman, Friedman, Grafman, & Perez, 1994; Chiarello, 1998; Grose-Fifer & Deacon, 2004) On the other hand, both hemispheres tend to show equivalent priming effects for more closely related or highly associated information (CAT-DOG) (Chiarello, 1998 #4639, but cf. Grose-Fifer & Deacon, 2004), as well as for repeated semantic information (Weems & Zaidel, 2005).

To summarize, the encoding literature has concentrated on hemispheric differences in processing of conceptual or perceptual form whereas hemispheric differences in semantic representation have been characterized as distinctions in breadth of activation. Evidence from both literatures suggests that the right and left hemisphere are involved in processing of conceptual verbal information, although each literature differs in their account of the kinds of information encoded or represented by each hemisphere.

How might the differences between hemispheric asymmetries in encoding and representation be accounted for? One possibility is that verbal information is subject to subsequent processing in each hemisphere after initial encoding. There has been little research examining the differences in the establishment of verbal representations across the hemispheres across time. To the extent it has been looked at such as the recent work by Federmeier and colleagues (Evans & Federmeier, 2007; Federmeier & Benjamin, 2005), it seems that the representation of verbal information may vary over time across the cerebral hemispheres. These studies find a LH advantage in retention over short time periods (from a few seconds), but this advantage disappears or is reversed over longer retention intervals (over a couple of minutes). It is possible that these asymmetries in retention result from an interaction between the information represented and encoded between the two hemispheres. At first, if the LH is preferentially encoding high level conceptual information about words, then this information may initially activate a broader semantic network that will promote subsequent identification of the same word, however, as this initial activation wanes and a more specific long-term representation of the word remains, this may result in slower response times to words, compared to the relatively broad/course and longer-lasting representation and/or activation of words in the RH.

In fact, priming studies that have examined activation of semantic information over relatively short time periods, as measured by stimulus onset asynchrony (SOA) of prime-target pairs ranging from 250-450 ms after prime presentation have provided support for this idea (Beeman, Friedman, Grafman, & Perez, 1994; Burgess & Simpson, 1988; Koivisto, 1997). These studies find that at short SOAs, the LH initially activates broad information that includes strongly and weakly related associates, but at longer SOAs this activation is

restricted to only stronger associates. On the other hand, the RH more slowly activates broader/coarse semantic information that remains active for longer durations.

Finally, despite wide evidence that the both cerebral hemispheres are involved in the encoding and representation of word meanings, there has been little research examining how these hemispheric differences may extend to the learning of novel word meanings. Developmental examinations of children with unilateral brain injury (Eisele & Aram, 1993; Thal et al., 1991),(Eisele & Aram, 1993; Thal et al., 1991) and ERP studies of early word knowledge in infants (Mills, Coffey-Corina, & Neville, 1997; Mills, Coffey-Corina, & Neville, 1993) suggest that both hemispheres are important in early word acquisition. Eisele and Aram (1993) measured single word naming and comprehension scores in children who had experienced LH or RH lesions at varying points during childhood and found lower scores in both groups on comprehension and naming compared to typically developing controls. However, children with RH lesions were relatively more impaired on comprehension measures than were those with LH lesions. Interestingly, the deficits in the RH group were not correlated with more general measures of cognitive performance like IQ, which suggest that these deficits were specific to language comprehension and naming. Similarly, Thal et al. (1991) measured early expressive and receptive vocabulary in a more uniform group of children who had experienced lesions at or around the time of birth via a parental report measure and found lower scores in children with both RH and LH lesions compared to typically developing controls, albeit with slightly greater impairment for those with RH damage. In addition, early ERP measurements of known and unknown word processing in children between the ages of 13 to 20 months of age have uncovered a shift in the scalp-recorded laterality of brainwave response to known words from bilateral at 13 months to

left-lateralized at 20 months, as well as increased left-lateralized responses in more children with higher vocabularies (Mills, Coffey-Corina, & Neville, 1997; Mills, Coffey-Corina, & Neville, 1993). This is consistent with a hypothesized change in the involvement of the two hemispheres in word processing from birth to 20 months. Altogether, these data suggest, that both hemispheres may mediate lexical learning in young children. Findings from lesion studies additionally suggest that the RH may have even greater involvement than the LH especially early in development.

Even though developmental work indirectly indicates involvement of both cerebral hemispheres in early lexical learning, there has been no work to directly assess this possibility in either children or adults. It is therefore the major goal of this study to examine the differential involvement of the two hemispheres in the acquisition of word meaning. We start by examining the hemispheric representation of novel word meanings in adults rather than in children for a variety of reasons. First, the primary method that has been established to probe hemispheric representation of word meaning, hemifield presentation, is more difficult to utilize with children, due to the methodological necessity that participants remain still and fixated on a central point as stimuli are rapidly flashed to the left (LVF) or right visual field (RVF). With adults, it is also possible to utilize visual written stimuli for lateralized presentation, whereas studies with children more often require pictures or auditory stimuli that are more difficult to accommodate in hemifield presentation techniques.

In addition, word learning by adults has become a topic of recent interest to electrophysiological researchers. Word learning in young adult's native language is commonplace and an important avenue of word learning. Estimates of the number of words known by adults vary widely, but generally fall between 40000 – 150000, whereas pre-

literate five or six year old children will know around 2500-13000 words (Aitchinson, 1994; Beck & McKeown, 1991; Bloom, 2000; Pinker, 1994). Clearly, adults and older children learn a tremendous number of words. The mode by which words are learned by school age children and adults tends to be almost entirely via incidental learning in various language contexts, especially during reading (Jenkins, Stein, & Wysocki, 1984; Nagy, Anderson, & Herman, 1987; Nagy, Herman, & Anderson, 1985; Sternberg, 1987). In addition, under the right conditions both children and adults can acquire knowledge of word meanings after only a single usage, a process called “fast-mapping” (Carey & Bartlett, 1978; Dollaghan, 1985; Heibeck & Markman, 1987). Moreover, recent electrophysiological research has indicated that the brain rapidly acquires relatively refined information about novel word meaning over the course of weeks (McLaughlin, Osterhout, & Kim, 2004) an hour (Perfetti, Wlotko, & Hart, 2005), several presentations in sentence (Mestres-Misse, Rodriguez-Fornells, & Munte, 2006) and even after just a single sentential presentation, in certain contexts (Borovsky, Elman, & Kutas, 2007, in prep). These studies that have examined acquisition after a single presentation in sentence contexts have shown that at least one contextual factor, sentential constraint, can significantly impact the likelihood of successfully integrating a novel word into one’s mental lexicon. In particular, sentential contexts that strongly constrain a novel word’s meaning by providing an expectation of a single, specific meaning can lead to significant understanding of a novel word’s meaning and usage after only a single exposure.

In this study, we use the visual half field technique in combination with a semantic priming paradigm to probe the two hemispheres after young adults are presented with known and unknown words in strongly and weakly constraining sentence contexts in central vision. In the priming task, both known and unknown words appear centrally as primes for

synonymous or unrelated target words that are laterally presented to either the RVF/LH or LVF/RH. In this way, we can examine the different processing consequences of giving each hemisphere a headstart on processing of a novel word recently experienced in a strongly or weakly constraining context.

5.3 Method

5.3.1 Participants

Participants were 24 healthy adults (17 women, 7 men; average age: 19.92, age range 18-24) and were given credit or paid \$7/hr for their participation. All participants were right-handed, native English speakers, and had no significant exposure to another language at least before the age of 12. Participants reported no history of mental illness, learning disability, language impairment, drug abuse, or neurological trauma. All participants had normal hearing and normal (or corrected to normal) vision. An additional 14 participated but were not analyzed: 6 had excessive blinking or motion artifact, 2 because of equipment failure or experimenter error, and 6 reported a characteristic which disqualified them from analysis (3 had significant second language exposure as a child, 2 had non-normal vision, 1 was not right-handed.)

5.3.2 Materials

Stimuli consisted of 128 sentence pairs selected from Federmeier and Kutas (1999), and 512 word pairs selected to correspond with 128 sentence final words. Both are described in detail below:

Table 5-1. Examples of sentences and word pairs in each condition

A) Context Sentences	
High/Known	Peter sat gaping at the centerfold. He asked his friend if he could borrow the MAGAZINE.
High/Unknown	Peter sat gaping at the centerfold. He asked his friend if he could borrow the YERGE.
Low/Known	The package was rectangular and heavy and suspiciously academic. Bianca was disappointed that her uncle was giving her a BOOK.
Low/Unknown	The package was rectangular and heavy and suspiciously academic. Bianca was disappointed that her uncle was giving her a SHUS.

B) Word Pairs			
	<u>Identical</u>	<u>Related</u>	<u>Unrelated</u>
High / Known	magazine - MAGAZINE	magazine - NOVEL	magazine - ACCIDENT
High / Unknown	yerge - MAGAZINE	yerge - NOVEL	yerge - ACCIDENT
Low / Known	book – BOOK	book - LETTER	book - ROAD
Low / Unknown	shus - BOOK	shus - LETTER	shus - ROAD

Note: all word pairs were also paired with an equal number of pseudoword targets, not depicted in this table

5.3.2.1 Sentences

64 high constraint and 64 low constraint sentence pairs were selected from Federmeier & Kutas (1999). These had been extensively normed to ensure adequate levels of cloze probability for high and low constraint sentences (Federmeier & Kutas, 1999). The Federmeier and Kutas (1999) sentence pairs consisted of an initial sentence that set up an expectation of a meaning and item category, and a second sentence that ended with words that were either plausible, expected word sentence completions, or unknown pseudowords. This yielded 32 sentences in each of four main conditions: 1) High constraint / Known word ending, 2) High Constraint / Unknown word ending, 3) Low constraint / Known word ending, and 4) Low constraint / Unknown word ending. Sentences were counterbalanced such that each High and Low constraint sentence pairs appeared with both a Known and Unknown

ending equally across all versions of the study, but not repeated within a subject. Known word target items consisted of words in 64 categories, and these target categories were used as the basis for selecting semantically related and unrelated prime-target pairs, described below. The sentence stimuli were counterbalanced across versions so that all sentences appeared with both Known and Unknown word endings with equal frequency across participants. Table 5-1a includes examples of sentence stimuli used.

5.3.2.2 Word-Pairs

512 word pairs were constructed for the study, consisting of a prime followed by a target word presented one stimulus at a time. Prime stimuli were unknown and known words from the sentence endings described above. Target stimuli for each prime were selected for two conditions: 1) Synonym/Identical meaning (rabbit-RABBIT) and 2) Unrelated (rabbit-RIBBON). Unrelated word pairs were selected to be as closely matched in number of letters and syllables, concreteness, imageability, familiarity and frequency, as reported by the MRC psycholinguistic database (Wilson, 1988). In cases involving Unknown word primes, the relatedness to the target (Related, Unrelated and Synonym) was determined by the implied meaning of the sentence context in which the Unknown word had previously appeared.. An equal number of Nonword targets were also constructed so that lexical decision responses should be approximately 50% “No”. Nonwords were constructed using the ARC Nonword database (Rastle, Harrington, & Coltheart, 2002), and were selected to be pronounceable, conform to English phonotactics, and contain between 4 and 7 letters. In any one version, participants saw each Known and Unknown prime paired with a three out of four possible real word targets and hemifield combinations. Versions were

counterbalanced such that each Known and Unknown prime was matched with the targets with equal frequency across versions. Table 5-1b includes examples of word pairs in the study. An equal number of nonword targets were also constructed.

5.3.3 Procedure

Participants were tested in a single experimental session conducted in a soundproof, electrically-shielded chamber and were seated in a comfortable chair in front of a monitor. Sessions consisted of two interleaved sections: sentence comprehension and priming.

In the sentence comprehension task, participants were instructed to read the context sentences for comprehension and to do their best to understand the sentence and words even when “nonsense” words appeared on the screen. The presentation of the first sentence in the sentence pair was presented in its entirety on the monitor, and participants were instructed to press a button to indicate that they had completed reading this sentence and were ready to see the second sentence. The second sentence was preceded by a series of crosses (500 ms duration with a stimulus-onset-asynchrony varying randomly between 300 and 800 ms) to orient the participant toward the center of the screen. Sentences were then presented one word at a time, each for 200 ms with a stimulus-onset-asynchrony of 500 milliseconds. Participants were asked to minimize blinking and movement as much as possible during sentence presentation. The final target word appeared on the screen for 1400 ms.

In the priming task, participants were instructed to read every word that appeared on the screen and indicate with a button press if the target items (which were identified by appearing in capital letters) were or were not real words. Participants viewed two sets of

prime/target pairs, and were given a 2500ms offset break period to blink between pairs. Prime pair onsets were preceded by a set of fixation crosses that were randomly presented for 200-500ms. A central fixation dot appeared on the screen throughout the trials positioned at a half degree immediately below the centrally presented prime words and fixation crosses. Immediately following the fixation cross, a prime word appeared centrally in lower case letters for 200 ms, followed by an offset of 300ms, followed by the uppercase target word that appeared in the left or right visual field with the inner edge 2° of visual angle from fixation for 200ms, and offset of 800ms. Participants provided a lexical decision response as soon as possible after the presentation of each target word in capital letters.

Sentence comprehension and priming tasks were interleaved as follows. Participants read 16 sentence pairs, and then completed the priming task consisting of 96 pairs, with primes being selected from the 16 sentences endings that had just been previously read. Participants were given a break before beginning a new block of sentences. The entire experiment consisted of 8 blocks of sentence/prime sets. At the end of the study, participants were asked to complete a surprise old/new recognition memory questionnaire containing 50 sentences that had appeared in the study, and 50 sentences that had not appeared. This was given to ensure that participants sufficiently attended to the sentences during the study.

5.3.4 Electrophysiological recording

Scalp potentials were continuously recorded from 26 geodesically arranged sites using an ElectroCap with tin electrodes. Electrodes were placed at equal distances across the scalp, with positions and labels shown in Figure 5-1 . A left mastoid reference was used.

Horizontal eye movements were monitored via electrodes placed on the outer canthus, with the left electrode as a reference. Blinks were monitored via electrodes placed on the infraorbital ridge of each eye, with each electrode referenced to the left mastoid. Potentials were digitized at a sampling rate of 250 Hz and hardware bandpass filter of 0.1-100Hz with Grass Amplifiers and stored on a hard disk for later analysis. Impedances were kept below 5 k Ω . The ERPs were stimulus-locked averages consisting of a 100-ms baseline and a 920 post-stimulus interval.

5.3.5 Data analysis

Data were re-referenced offline to an average left and right mastoid. Trials contaminated by horizontal and vertical eye movements, blinks, excessive muscle activity, or amplifier blocking were rejected off-line before averaging. This process ensured that trials were only analyzed in which participants were centrally fixated and remaining still. ERPs were computed for epochs extending from 100 ms before stimulus onset to 920 ms after stimulus onset. Averages of ERP trials that were free of artifacts (including blinks, and horizontal eye movement) were computed to targets in all priming conditions in both the RVF/LH and LVF/RH: Identical/Synonym and Unrelated targets for each of the four main conditions High/Known, High/Unknown, Low/Known, Low/Unknown after subtraction of the 100 ms pre-stimulus baseline.

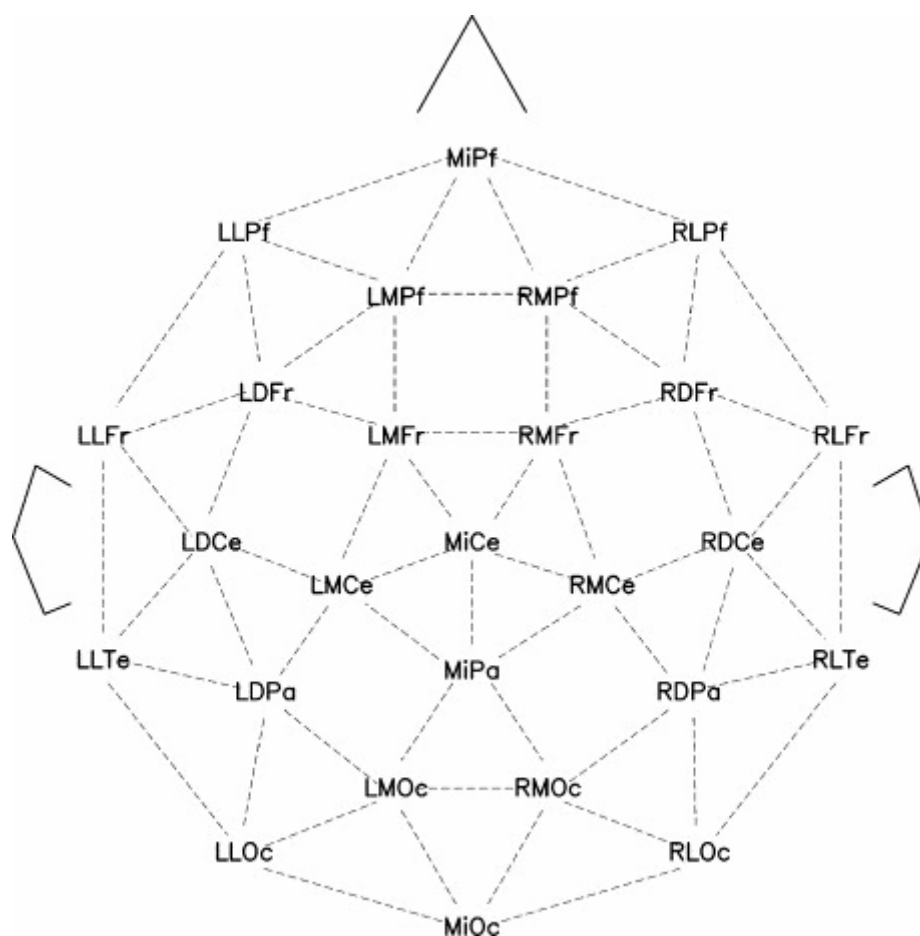


Figure 5-1. Diagram of electrode positions and labels

5.4 Results

5.4.1 Behavioral responses

5.4.1.1 Accuracy

Mean accuracy scores are shown in Figure 5-2. A four-factor repeated measures ANOVA with factors of Word type (Known and Unknown), Constraint (Low and High), Visual Field (LVF and RVF) and Prime relationship (Synonym/ID and Unrelated) revealed main effects of Word Type

[$F(1,23)=34.68$, $p<0.0001$], with lower accuracies for targets preceded by Unknown than Known words, Visual Field [$F(1,23)=53.55$, $p<0.0001$], with higher accuracies for words presented to the RVF/LH than LVF/RH, and Prime relationship [$F(1,23)=169.89$, $p<0.0001$], with higher accuracies observed when targets were preceded by a word that was related than unrelated in meaning. These main effects were modulated by interactions of Word Type x Visual Field [$F(1,23)=14.52$, $p=0.0009$], Word Type x Prime Relationship [$F(1,23)=100.19$, $p<0.0001$] and Word Type x Prime Relationship x Visual Field [$F(1,23)=10.53$, $p=0.0036$]. Marginal interactions of Visual Field x Prime [$F(1,23)=4.18$, $p=.0524$] and Word Type x Constraint x Prime [$F(1,23)=3.88$, $p=0.0611$] were also observed. Follow-up Tukey-test comparisons revealed that there were more correct responses to targets when presented to the RVF (than LVF), for both Known and Unknown word primes, and that Targets that were related to Known and Unknown word prime meanings also elicited higher accuracies (all $p < 0.05$). The three-way Word Type x Constraint x Prime interaction was driven by a lack of priming effect for Unknown words that initially appeared in Low constraint context whereas all other prime word conditions (Unknown/High, Known/High and Known/Low) showed reduced accuracies to targets that were unrelated in meaning (all $p < 0.05$). Additional Tukey tests also revealed that the three-way Word Type x Visual Field x Prime interaction was driven by an absence of a priming effect for Unknown words in the LVF/RH, while all other conditions (Known/RVF, Known/LVF, Unknown/RVF) showed reduced accuracies to unrelated target words (all $p < 0.05$). In sum, higher accuracies were observed for words that appeared in the RVF/LH compared to the LVF/RH, and for targets preceded by Known, compared to Unknown words. In general, accuracies were reduced when primed by

words that were unrelated in meaning, except when preceded by Unknown words that initially appeared in low constraint contexts.

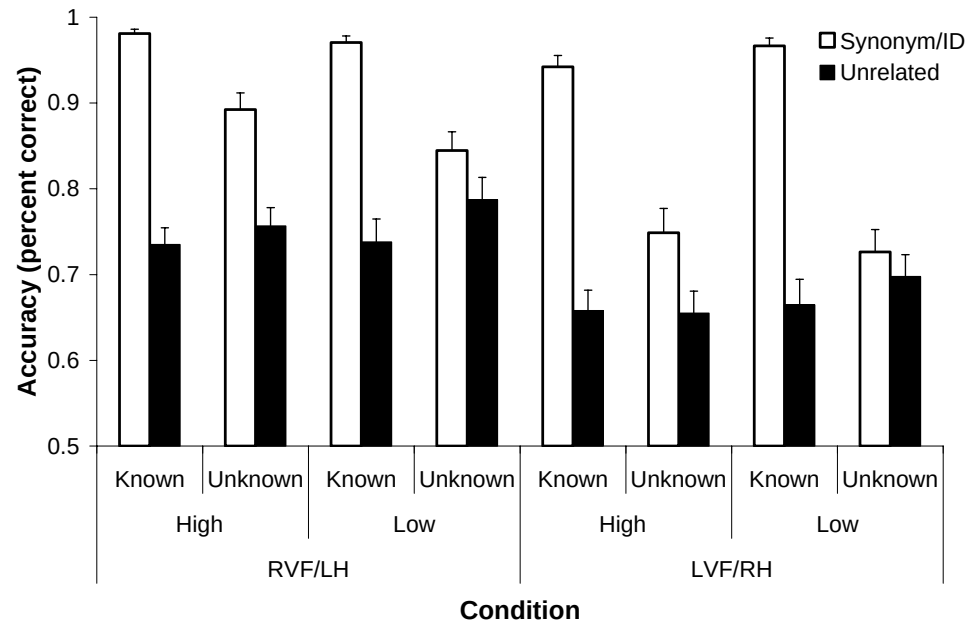


Figure 5-2. Mean Accuracy of behavioral responses across experimental conditions.

5.4.1.2 Reaction Time

Mean reaction times for correct responses are shown in Figure 5-3. A four-factor repeated measures ANOVA was carried out with factors of Word type (Known and Unknown), Constraint (Low and High), Visual Field (LVF and RVF) and Prime relationship (Synonym/ID and Unrelated). This analysis revealed main effects of Word Type [$F(1,23)=136.66$, $p<0.0001$], with faster RTs for targets preceded by Known words, Visual Field [$F(1,23)=26.59$, $p<0.0001$], with faster RTs for targets presented in the RVF/LH, and Prime [$F(1,23)=170.44$, $p<0.0001$], with slower RTs to targets preceded by Unrelated words. Interactions of Word Type x Prime [$F(1,23)=96.01$, $p<0.0001$], Visual Field x Prime

[$F(1,23)=7.36$, $p=0.0124$], Word Type x Constraint x Prime [$F(1,23)=5.53$, $p=0.0277$], Word Type x Constraint x Visual Field [$F(1,23)=7.38$, $p=0.0123$]. A marginal effect of Constraint x Visual Field was also observed [$F(1,23)=3.98$, $p=0.0581$] Follow-up Tukey test comparisons revealed that these interactions were driven by a lack of priming effect for targets that were preceded by Unknown words. In addition, targets preceded by both Unknown/High and Unknown/Low words did not show priming effects. Significant effects of priming were observed in both visual fields, and as expected, were observed for targets preceded by Known words that had appeared in both High and Low constraint contexts (all $p < 0.05$). In sum, robust priming effects were found for Known words that had initially appeared in sentences of both High and Low constraint, however, priming effects were not found for targets preceded by Unknown words.

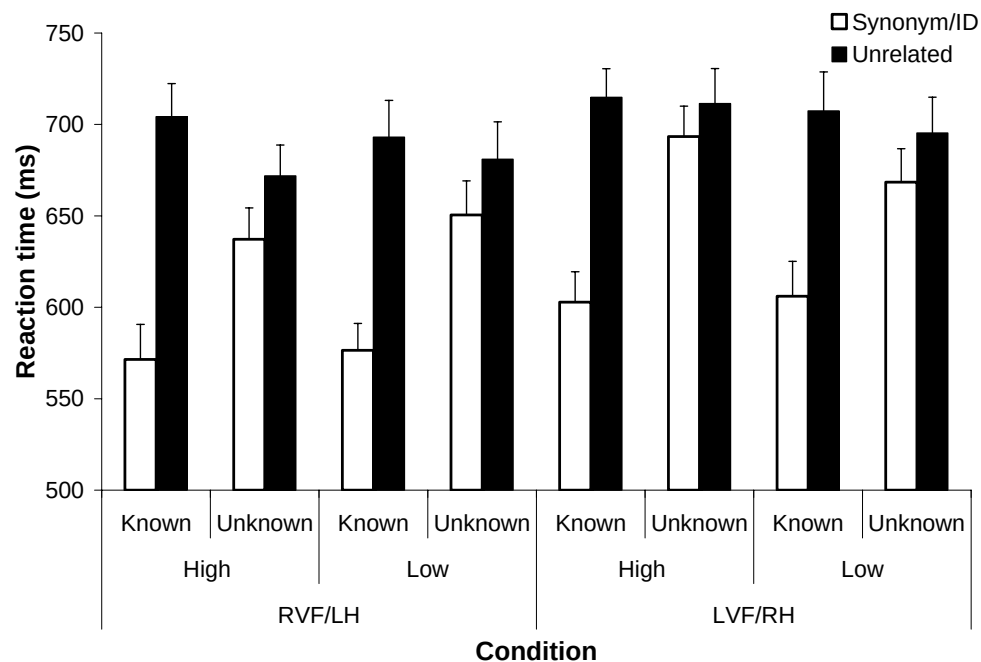


Figure 5-3. Mean reaction time of behavioral responses across experimental conditions

In order to further examine the priming relationship across all conditions, we also carried out items analyses with a four factor ($2 \times 2 \times 2 \times 2$) repeated measures ANOVA with identical factors to the by-subjects analysis carried out above. Like the subjects analysis, these tests also revealed main effects of Word Type [$F(1,1981)=20.07$, $p<0.0001$], with targets preceded by Known words eliciting faster responses, a main effect of Prime Relationship [$F(1,1981)=142.65$, $p<0.0001$], with targets preceded by primes that are Identical in meaning eliciting faster reaction times, and a main effect of Visual Field [$F(1,1981)=17.89$, $p<0.0001$]. A two-way interaction of Word Type x Prime Relationship [$F(1,1981)=41.89$, $p<0.0001$] was also observed, with follow-up Tukey tests indicating that this interaction was driven significant differences between every Word Type and Prime relationship pair, except for targets preceded by Unrelated Known and Unknown words. No other significant interactions were observed. Although no significant four-way interaction was observed, Tukey tests pairwise comparisons were conducted to examine the relationships between Identical, Related and Unrelated meanings in each of the four prime conditions: Known/High, Unknown/High, Known/Low, and Unknown/Low across both Visual Fields. These analyses revealed that targets in both visual fields preceded by Known/High and Known/Low word primes elicited faster reaction times when preceded by a word identical in meaning, compared to an unrelated word. On the other hand, targets preceded by Unknown words did not elicit priming effects in any condition.

5.4.2 Electrophysiological responses

Artifact free correct responses to target words in eight priming conditions were analyzed, four each in the RVF/LH and the LVF/RH conditions: Known/Related,

Known/Unrelated, Unknown/Related, Unknown/Unrelated. Grand average ERPs across all electrodes to target words in all eight conditions are shown in Figure 5-5 through Figure 5-8 and in at a single electrode, the vertex, in Figure 5-9. As can be seen from these figures, an effect of target type is seen in all four Known word priming conditions, and in one Unknown word priming condition, represented by a modulation of a negative going peak ranging between 250-500ms (N400), followed by a continuing positivity between 500-900ms (LPC). We analyzed the ERP across three time windows corresponding to the N1 component (100-200ms), N400 (250-500ms), and LPC (500-900ms).

5.4.2.1 N100 amplitude

In order to ensure that initial processing of the target words was conducted solely by the contralateral hemisphere, we examined the N1, which has been shown to reflect initial perception and awareness of visual stimuli. Mean amplitude N1 responses were measured across 10 posterior electrodes (5 in each hemisphere) where N1 effects are typically largest in a time window between 100-200ms to correct responses in all experimental conditions. The Left hemisphere electrodes were: LLOc, LLTe, LMOc, LDPa and LDCe, and corresponding Right hemisphere electrodes were: RLOc, RLTe, RMOc, RDPa, and RDCe.

Figure 5-4 shows the effect of visual field on presentation at LLOc and RLOc. A two-factor repeated measures ANOVA was conducted with factors of VF (LVF and RVF), and Hemisphere (measured at RH or LH electrodes). This analysis revealed main effects of Hemisphere [$F(1,23)=28.48$, $p<0.0001$] and VF [$F(1,23)=6.84$, $p<0.0089$]. An interaction of Hemisphere and VF was also significant [$F(1,23) =153.10$, $p<0.0001$]. Follow-up Tukey tests revealed that N1 amplitudes were larger (more negative) over the contralateral hemisphere

of visual presentation (mean amplitude for RVF in LH: $1.58\mu\text{V}$ and RH: $2.16\mu\text{V}$, and for LVF presentation in LH: $2.38\mu\text{V}$ and RH: $0.92\mu\text{V}$), and that the largest N1 amplitudes in each hemisphere were evoked by stimuli from the contralateral hemisphere (mean N1 amplitude over LH in RVF: $1.58\mu\text{V}$ and LVF: $2.38\mu\text{V}$ and in over RH in RVF: $2.16\mu\text{V}$ and LVF: $0.92\mu\text{V}$).

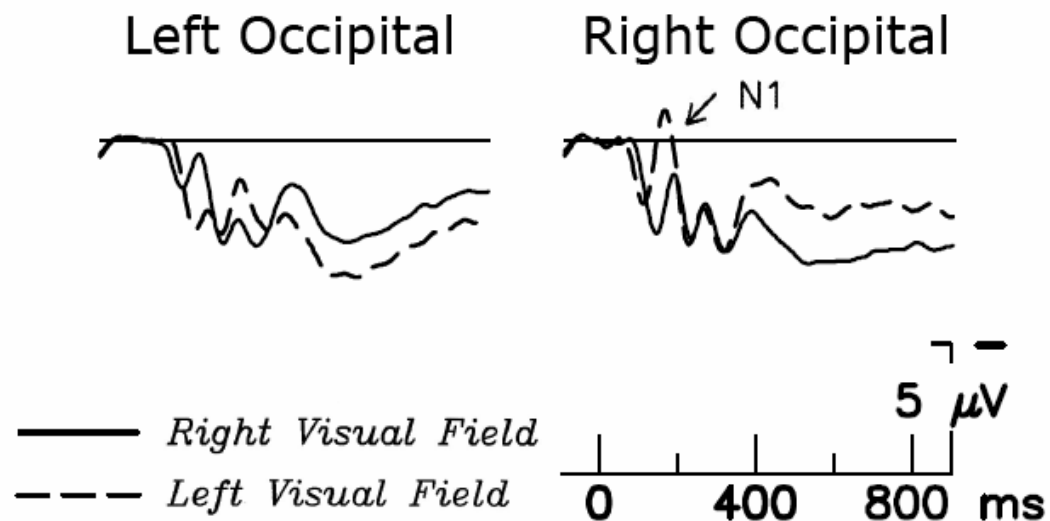


Figure 5-4. Effect of stimulus lateralization on the ERP. ERPs are plotted across two electrodes where N1 effects of typically largest as a function of visual field of stimulus presentation, collapsed across constraint, prime relationship, and word type. N1 responses (100-200ms) are largest in the contralateral field of stimulus presentation.

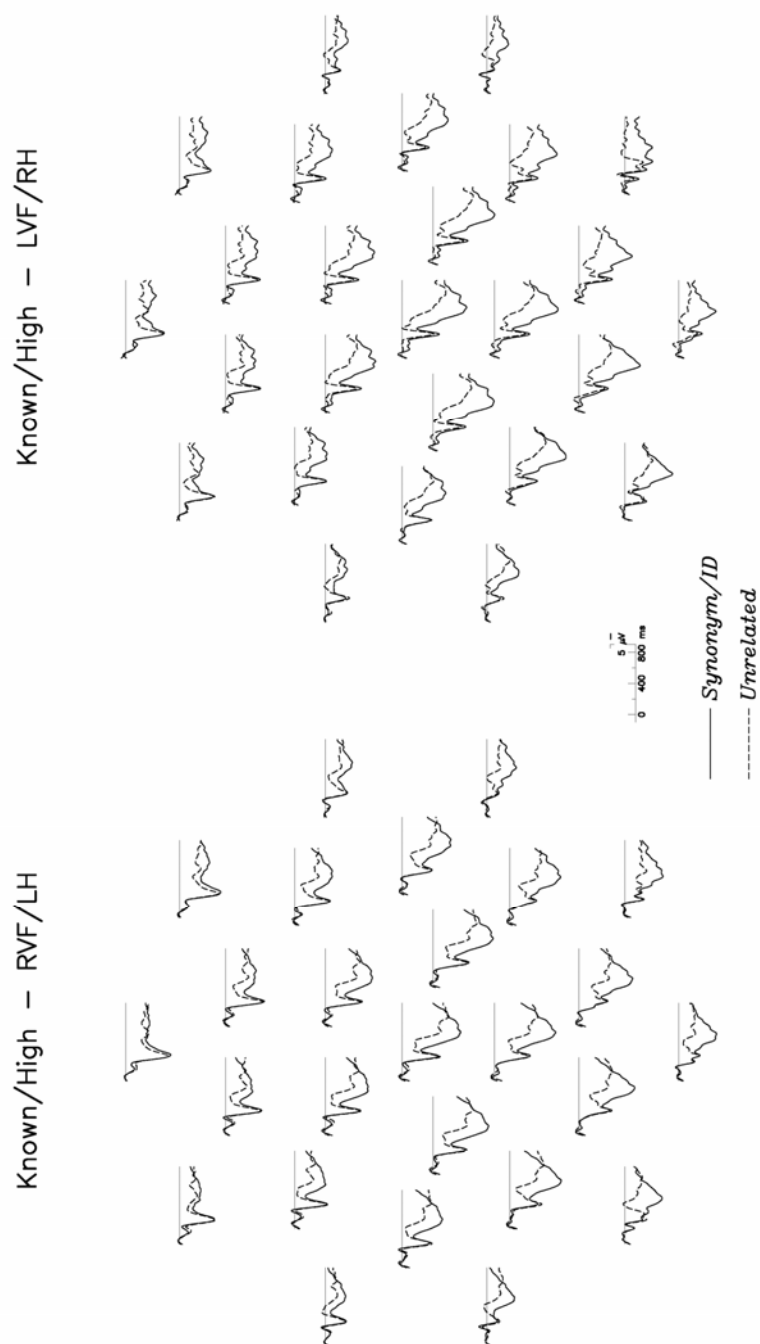


Figure 5-5. Grand average ERPs across all subjects and electrode sites for targets presented to RVF/LH and LVF/RH that were preceded by Known primes presented in High constraint sentence contexts.

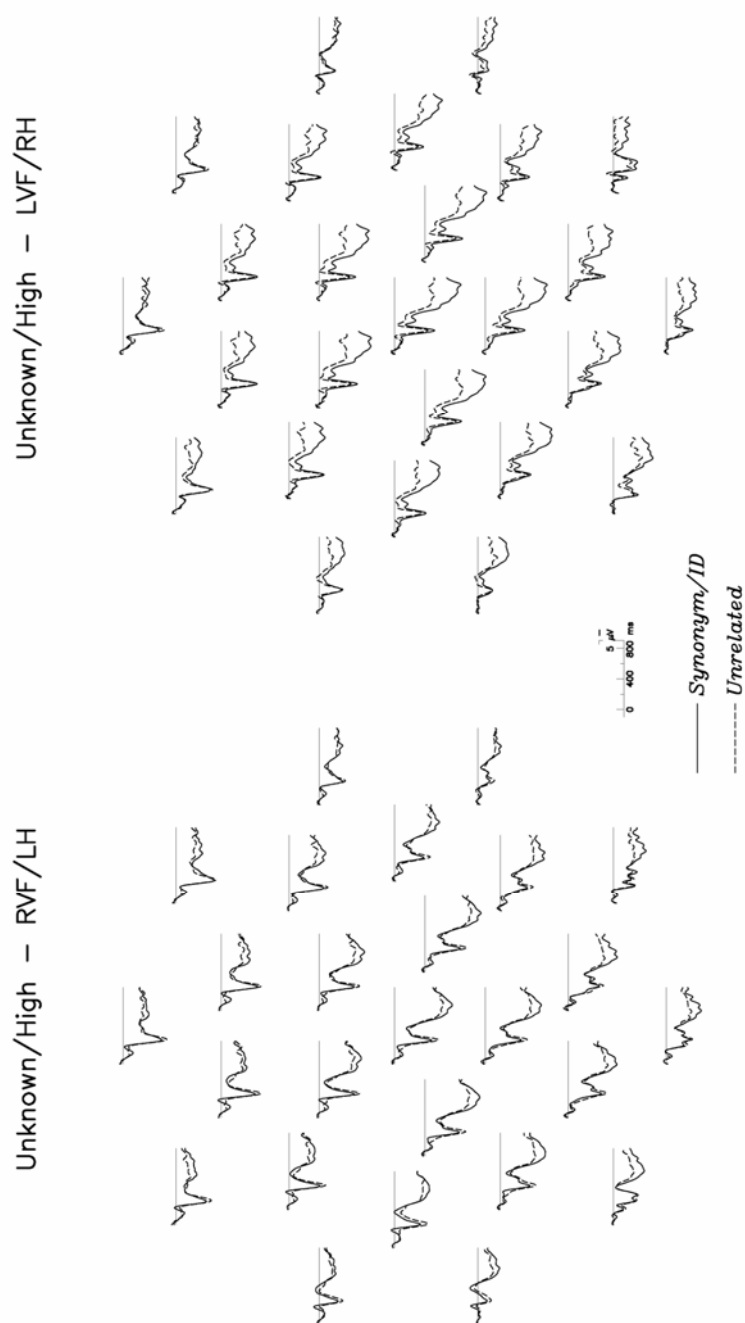


Figure 5-6. Grand average ERPs across all subjects and electrode sites for targets presented to RVF/LH and LVF/RH that were preceded by Unknown primes presented in High constraint sentence contexts

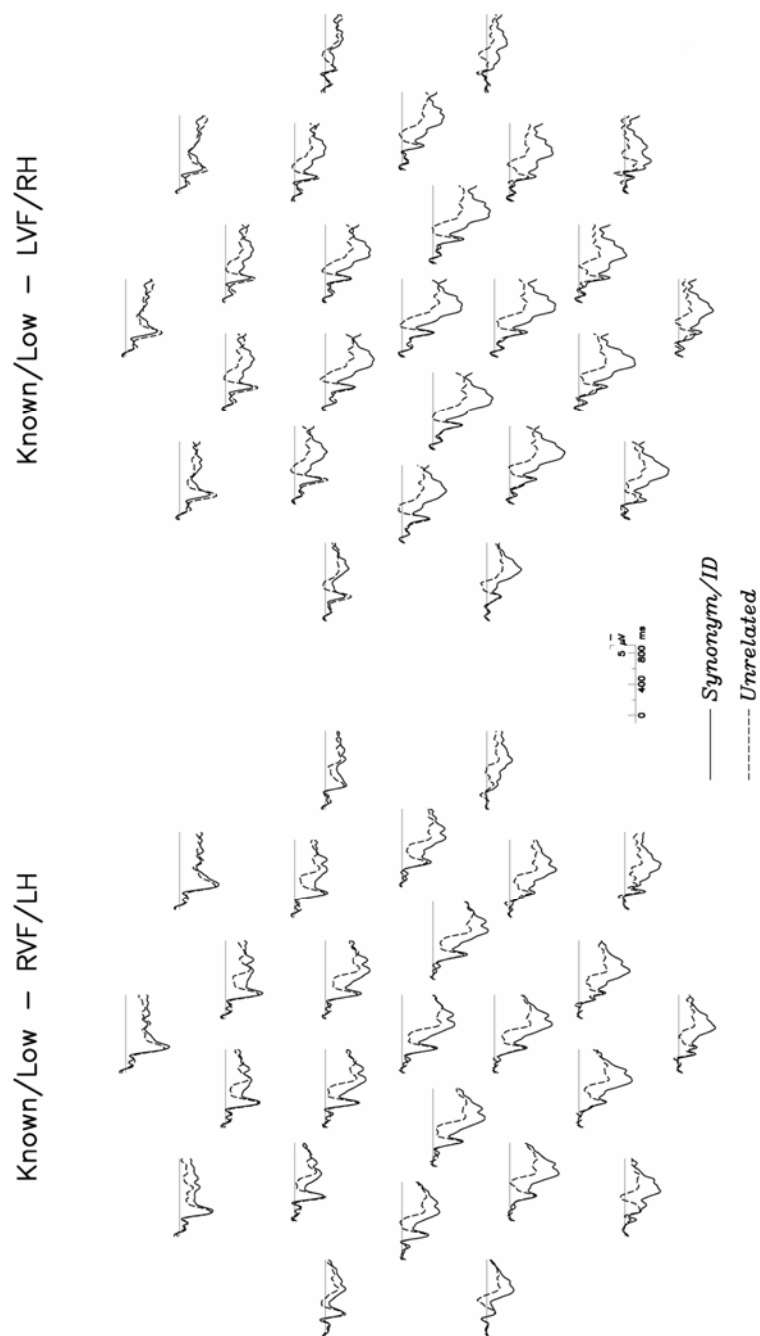


Figure 5-7. Grand average ERPs across all subjects and electrode sites for targets presented to RVF/LH and LVF/RH that were preceded by Known primes presented in Low constraint sentence contexts

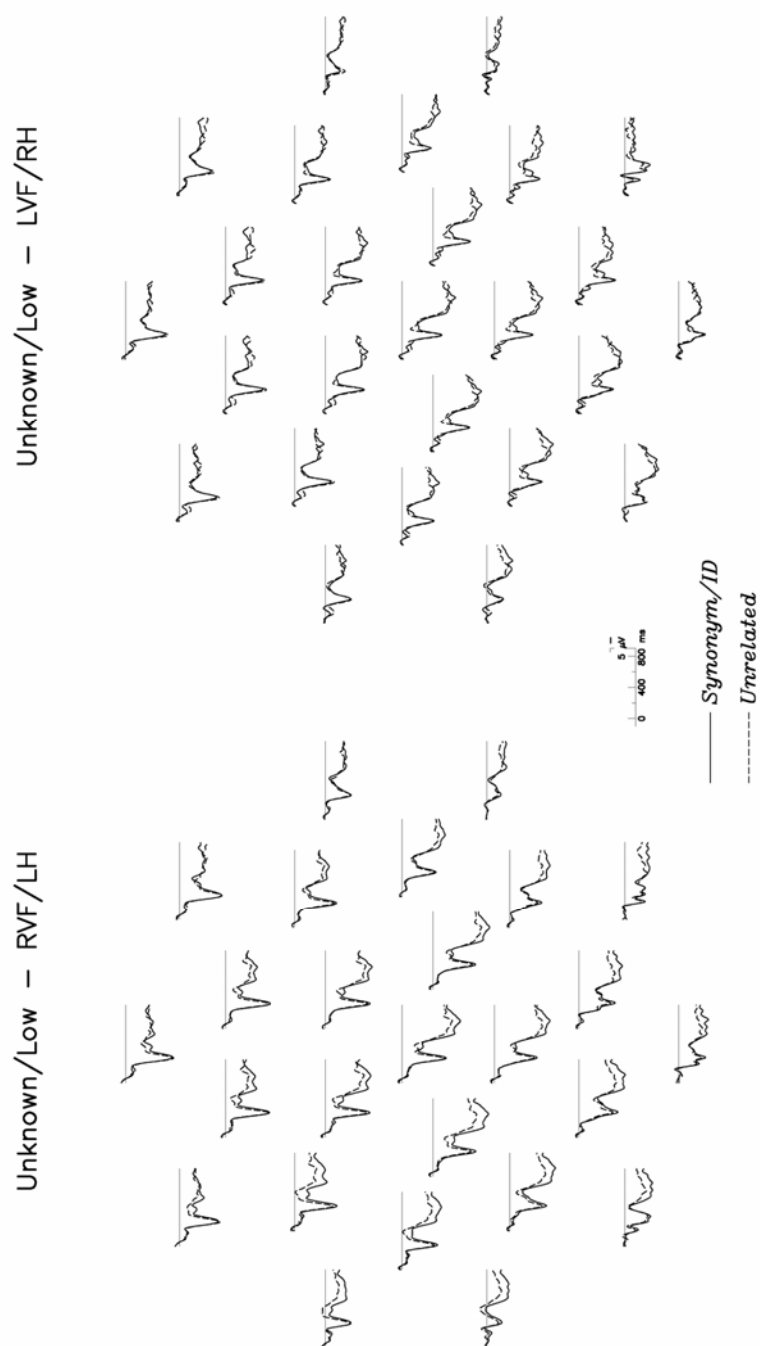


Figure 5-8. Grand average ERPs across all subjects and electrode sites for targets presented to RVF/LH and LVF/RH that were preceded by Unknown primes presented in Low constraint sentence contexts

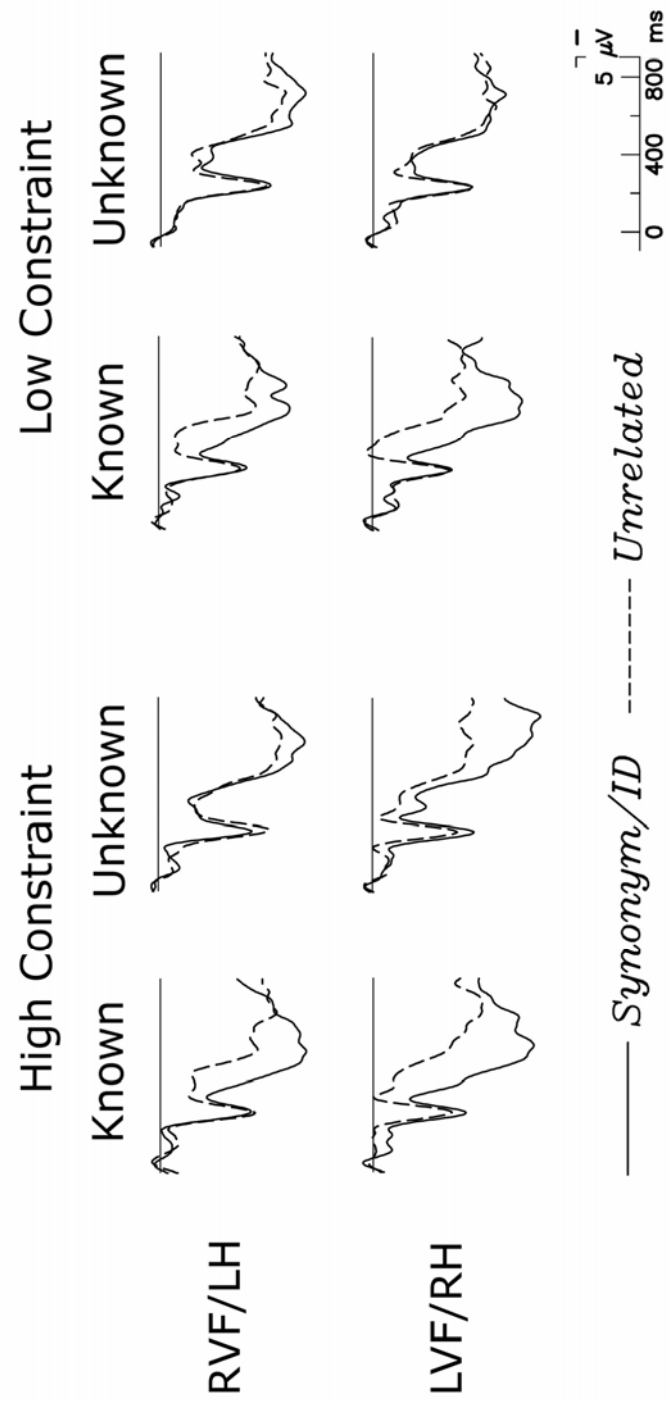


Figure 5-9. Grand average ERPs to target words in priming task at a single electrode, MiCE

5.4.2.2 N400 amplitude

In order to examine N400 effects across experimental conditions, mean N400 amplitudes were measured in a time window between 250-500ms across all 26 channels to correct responses. A repeated measures ANOVA was conducted with factors of Word Type (Known and Unknown) x Constraint (High and Low) x Visual Field (LVF/RH, and RVF/LH), Prime Relatedness (Synonym/ID and Unrelated) and Channel (26 levels of electrode). This analysis revealed main effects of all model factors: [Word Type: $F(1,23)=73.59$, $p<0.0001$, Constraint: $F(1,23)=4.9381$, $p<0.0263$, Prime: $F(1,20)=1267.114$, $p<0.0001$, Visual Field: $F(1,23)=141.89$, $p<0.0001$, and Channel: $F(25, 575)=56.53$, $p<0.0001$]. Six two-way interactions were also observed between Prime x Word type [$F(1,23)=536.57$, $p<0.0001$], Prime x Visual Field [$F(1,23)=16.77$], Constraint x Visual Field [$F(1,23)=4.26$, $p<0.0001$], Word Type x Constraint [$F(1,23)=40.66$, $p<0.0001$], Visual Field x Channel [$F(25,575)=3.3476$, $p<0.0001$], and Prime x Channel [$F(25,575)=5.5271$, $p<0.0001$]. Three-way interactions were observed between Prime x Constraint x Visual Field [$F(1,23)=12.743$, $p<0.0004$], Prime x Word type x Constraint [$F(1,23)=7.23$, $p<0.0072$] and Prime x Word Type x Channel [$F(25,575)=3.0876$, $p<0.0001$]. A four-way interaction between Prime x Word Type x Constraint x Visual Field was also found [$F(1,23)=6.9877$, $p<0.0001$]. In order to better understand the priming effects across Word Type, Constraint and Visual Field, eight planned comparisons were conducted that contrasted mean N400 priming effects for Synonym/ID and Unrelated targets in eight conditions: 1) Known/High/RVF, 2) Known/High/LVF, 3)Known/Low/RVF, 4) Known/Low/LVF, 5)Unknown/High/RVF, 6)Unknown/High/LVF, 7)Unknown/Low/RVF, and 8)Unknown/Low/LVF. As expected, significant priming effects were observed for all Known word conditions: Known/High/RVF [$F(1,23)=28.49$, $p<0.0001$],

Known/High/LVF [$F(1,23)=62.84$, $p<0.0001$], Known/Low/RVF [$F(1,23)=16.57$, $p<0.0005$], Known/Low/LVF [$F(1,23)=25.29$, $p<0.0001$]. In addition, a significant priming effect was observed for targets that were preceded by Unknown/High primes in the LVF [$F(1,23)=12.64$, $p<0.0017$]. No other priming effects in the N400 time window were observed.

5.4.2.3 N400 distribution

In order to examine if the priming effects observed above show equivalent distributions across the scalp, additional analyses were conducted to measure the distribution of the N400 amplitude in each condition. Figure 5-10 depicts the distribution of the N400 priming effect. These analyses were conducted over the N400 time-window (250-500ms) across 16 electrodes classified according to their position on the scalp. These distributional factors were: Hemisphere (Right or Left), Anteriority (Prefrontal, Frontal, Central, Occipital), Laterality (Medial or Lateral). The electrodes used in this analysis were: LLPf, RLPf, LMPf, RMPf, LLFr, RLFr, LMFr, RMFr, LLTe, RLTe, LMCE, RMCE, LLOc, RLOc, LMOc, RMOc (see Federmeier Mai & Kutas, 2005). A repeated measures ANOVA was conducted with experimental factors of Word Type (Known and Unknown) x Constraint (High and Low) x Visual Field (LVF/RH, and RVF/LH), and Prime Relatedness (Synonym/ID and Unrelated), and distributional electrode factors of Hemisphere (Right and Left), Anteriority (Prefrontal, Frontal, Central, Occipital) and Laterality (Medial and Lateral). Main effects were observed for the following factors: Word Type [$F(1,23)=42.08$, $p<0.0001$], Prime Relatedness [$F(1,23)=703.22$, $p<0.0001$], Visual Field [$F(1,23)=81.98$, $p<0.0001$], Hemisphere [$F(1,23)=9.80$, $p=0.0018$], Anteriority [$F(3, 69)=69.38$, $p<0.0001$], and Laterality [$F(1,23)=483.95$, $p<0.0001$]. Interactions of distributional factors were found for Hemisphere x Anteriority [$F(3,69)=3.77$,

$p=0.0102$], and Anteriority $\times$ Laterality [$F(3,69)=109.36$, $p<0.0001$]. Main effects and interactions of distributional factors were driven by a tendency for the N400 to be larger in the right hemisphere, over central and lateral scalp locations. There was an interaction of Visual Field $\times$ Hemisphere, [$F(1,23)=48.68$, $p<0.0001$], that was driven by larger N400 amplitudes to stimuli presented in the LVF/RH over RH electrode sites, compared to stimuli presented to the RVF/LH over the RH electrodes. There were also an additional interaction of Visual Field $\times$ Hemisphere $\times$ Laterality [$F(1,23)=12.73$, $p=0.0004$] that was driven by larger N400 amplitudes for stimuli presented to the LVF/RH, except at left lateral electrodes. Additional interactions were observed for: Prime $\times$ Word type [$F(1,23)=291.70$, $p<0.0001$], Word Type $\times$ Constraint [$F(1,23)=21.13$, $p<0.0001$], Prime $\times$ Visual Field [$F(1,23)=7.87$, $p=0.0051$], Prime $\times$ Constraint $\times$ Visual Field [$F(1,23)=7.49$, $p=0.0062$], Prime $\times$ Anteriority [$F(3,69)=10.96$, $p<0.0001$], Prime $\times$ Word Type $\times$ Anteriority [$F(3,69)=8.49$, $p<0.0001$], Prime $\times$ Laterality [$F(1,23)=54.96$, $p<0.0001$], Prime $\times$ Word Type $\times$ Laterality [$F(1,23)=21.57$, $p<0.0001$], Prime $\times$ Visual Field $\times$ Laterality [$F(1,23)=5.15$, $p=0.0233$], Visual Field $\times$ Hemisphere $\times$ Laterality [$F(1,23)=12.73$, $p=0.0004$]. Despite there being a significant four-way interaction between experimental factors (as observed in the amplitude analyses above) [$F(1,23)=3.99$, $p=0.046$], no interactions were found between this four-way combination of experimental factors and any individual distributional factors, or combination of distributional factors, indicating that experimental priming effects did not have differential distributions. In sum, the N400 effects across analyses over the restricted and larger set of electrodes reveal priming effects for the Known word conditions, and the Unknown/High condition, but not the Unknown/Low. In addition, the N400 priming effects observed in the Unknown/High condition did not differ in distribution from the Known word priming effects.

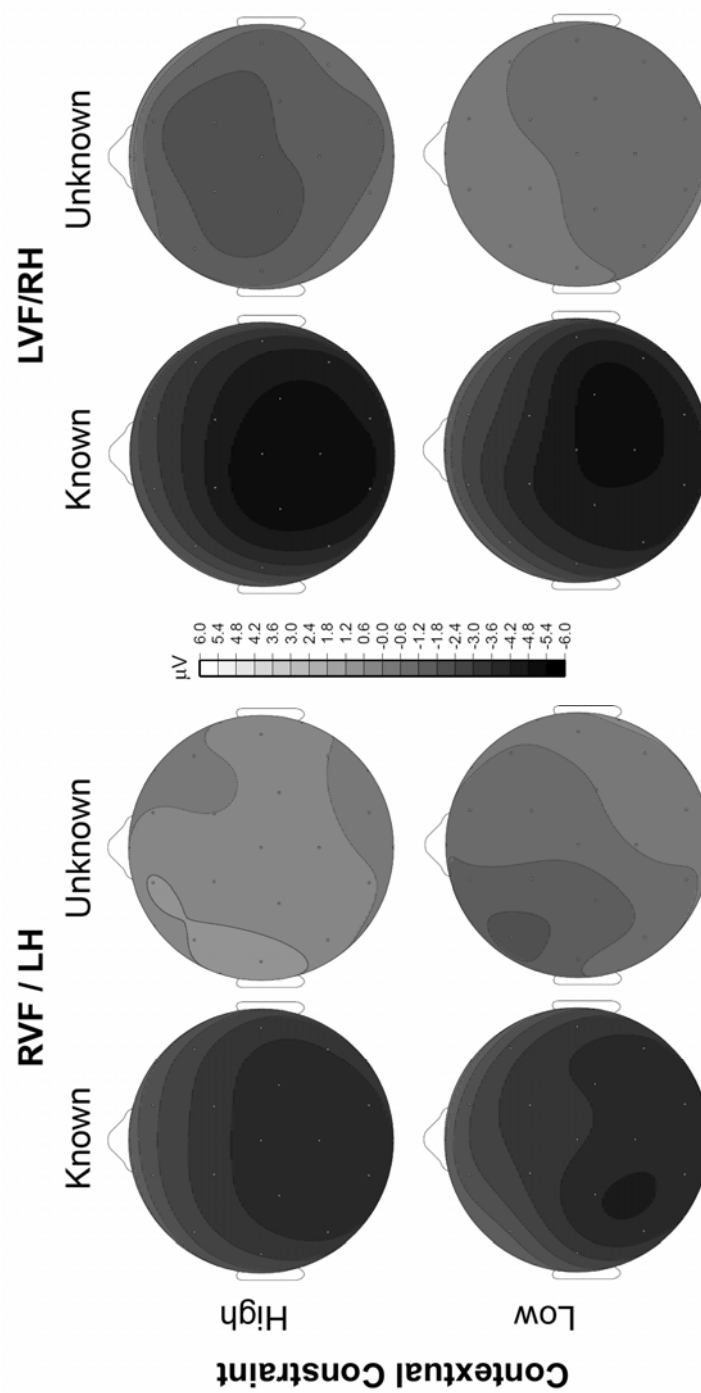


Figure 5-10. Spline interpolated isovoltage maps plot the distribution of the priming effect (Unrelated – Synonym/ID) from 250-500ms across experimental conditions in both visual fields.

5.4.2.4 LPC

In addition to experimental effects observed in the traditional N400 time window between 250-500ms, there were extended priming effects that continued through the 900ms epoch. In order to examine these late effects, we calculated mean LPC amplitude from a period between 500-900ms across all 26 electrodes for correct responses. A repeated measures ANOVA was conducted with factors of Word Type (Known and Unknown) x Constraint (High and Low) x Visual Field (LVF/RH, and RVF/LH), Prime Relatedness (Synonym/ID and Unrelated) and Channel (26 levels of electrode). Main effects of Prime [$F(1,23)=623.92$, $p<0.0001$], Word Type [$F(1,23)=67.25$, $p<0.0001$], Constraint [$F(1,23)=30.98$, $p<0.0001$], Visual Field [$F(1,23)=14.36$, $p<0.0001$] and Channel [$F(25,575)=147.63$, $p<0.0001$] were observed. Significant two-way interactions included Prime x Constraint [$F(1,23)=25.68$, $p<0.0001$], Prime x Channel [$F(25,575)=3.36$, $p<0.0001$], Prime x Visual Field [$F(1,23)=18.74$, $p<0.0001$], Word Type x Visual Field [$F(1,23)=$] and Visual Field x Channel [$F(25,575)=6.64$, $p<0.0001$]. Three way interactions were: Prime x Word Type x Constraint [$F(1,23)=4.21$, $p<0.0402$], Prime x Word Type x Visual Field [$F(1,23)=4.287$, $p<0.0385$], Prime x Constraint x Visual Field [$F(1,23)=33.25$, $p<0.0001$]. A four-way Prime x Constraint a Word Type x Visual Field interaction was also observed [$F(1,23)=18.16$, $p<0.0001$]. As with the N400 analysis Follow-up comparisons were conducted in order quantify the extent of late priming effects across experimental conditions. Synonym/ID and Unrelated LPC mean amplitude was compared across eight experimental conditions: 1)Known/High/RVF, 2) Known/High/LVF, 3)Known/Low/RVF, 4) Known/Low/LVF, 5)Unknown/High/RVF, 6)Unknown/High/LVF, 7)Unknown/Low/RVF, and 8)Unknown/Low/LVF. Significant LPC priming effects were found for all eight conditions but one: Unknown/Low/LVF. F-values for the remaining seven

significant comparisons were: Known/High/RVF: $[F(1,23)=7.58, p<0.0113]$, Known/High/LVF: $[F(1,23)=28.95, p<0.0001]$, Known/Low/RVF: $[F(1,23)=5.97, p<0.0226]$, Known/Low/LVF: $[F(1,23)=11.89, p<0.0022]$, Unknown/High/RVF: $[F(1,23)=8.46, p<0.0079]$, Unknown/High/LVF: $[F(1,23)=24.78, p<0.0001]$, Unknown/Low/RVF: $[F(1,23)=8.99, p<0.0064]$.

5.4.2.5 LPC Distribution

In order to characterize the distribution of the LPC effect, additional analyses were carried out within the LPC time window (between 500-900ms) over 16 electrodes, classified according to scalp location, as described in the N400 distributional analyses above. Figure 5-11 shows the distribution of the LPC priming effect across the scalp. A repeated measures ANOVA was conducted with experimental factors of Word Type (Known and Unknown) x Constraint (High and Low) x Visual Field (LVF/RH, and RVF/LH), and Prime Relatedness (Synonym/ID and Unrelated), and distributional electrode factors of Hemisphere (Right and Left) x Anteriority (Prefrontal, Frontal, Central, Occipital) x Laterality (Medial and Lateral). Main effects were observed for the following factors: Word Type $[F(1,23)=35.88, p<0.0001]$, Constraint $[F(1,23)=16.07, p<0.0001]$, Prime Relatedness $[F(1,23)=340.86, p<0.0001]$, Visual Field $[F(1,23)=6.58, p=0.0103]$, Anteriority $[F(3,69)=97.66, p<0.0001]$, and Laterality $[F(1,23)=1464.21, p<0.0001]$. These main effects were driven by reduced LPC for targets preceded by words that were Known, Unrelated and Low constraint, compared to Unknown, Synonym/ID and High constraint, respectively. Targets that were presented to the LVF/RH also showed reduced LPC amplitude, compared to RVF/LH. Distributional main effects were driven by a tendency for the LPC to be largest over Left/Medial Frontal, Central and Posterior electrode sites, and smaller at Prefrontal and Lateral regions on the scalp. Interaction of

Visual Field by Hemisphere [$F(1,23)=76.30$, $p<0.0001$], Hemisphere x Anteriority [$F(3,69)=751.07$, $p<0.0001$], Visual Field x Anteriority [$F(3,69)=3.8$, $p=0.0265$] and Visual Field x Hemisphere x Anteriority [$F(3,69)=9.03$, $p<0.0001$], Visual Field x Hemisphere x Laterality [$F(1,23)=14.21$, $p=0.0002$], Visual Field x Hemisphere x Anteriority x Laterality [$F(3,69)=2.77$, $p=0.0402$] that was driven by larger (more positive) LPC amplitudes in electrodes over the medial electrodes in contralateral hemisphere of the visual field of the stimulus, and for stimuli presented to the RVF/LH to show slightly larger LPC amplitudes at posterior sites compared to LVF/RH stimuli. Other two- and three-way interactions were observed: Prime x Constraint [$F(1,23)=15.74$, $p<0.0001$], Prime x Word Type [$F(1,23)=11.69$, $p=0.0006$], Prime x Visual Field [$F(1,23)=7.90$, $p=.005$], Word Type x Visual Field [$F(1,23)=3.96$, $p=0.0468$], Prime x Constraint x Visual Field [$F(1,23)=20.19$, $p<0.0001$], Prime x Anteriority [$F(3,69)=315.38$, $p<0.0001$], Prime x Laterality [$F(1,23)=17.26$, $p<0.0001$], Word type x Laterality [$F(1,23)=6.48$, $p=0.0109$], Prime x Visual Field x Laterality [$F(1,23)=7.02$, $p=0.0081$], Hemisphere x Laterality [$F(1,23)=3.85$, $p=0.0499$], Anteriority x Laterality [$F(3,69)=145.38$, $p<0.0001$], Hemisphere x Anteriority x Laterality [$F(3,69)=6.48$, $p=0.0002$], Like in the previous analysis, a four-way interaction between Prime Relatedness, Word Type, Constraint and Visual Field was observed [$F(1,23)=10.15$, $p=0.0014$], but there were no further interactions between these four experimental factors and any single distributional factor or combination of distributional factors, suggesting that the LPC distribution of the priming effects observed was equivalent across experimental conditions.

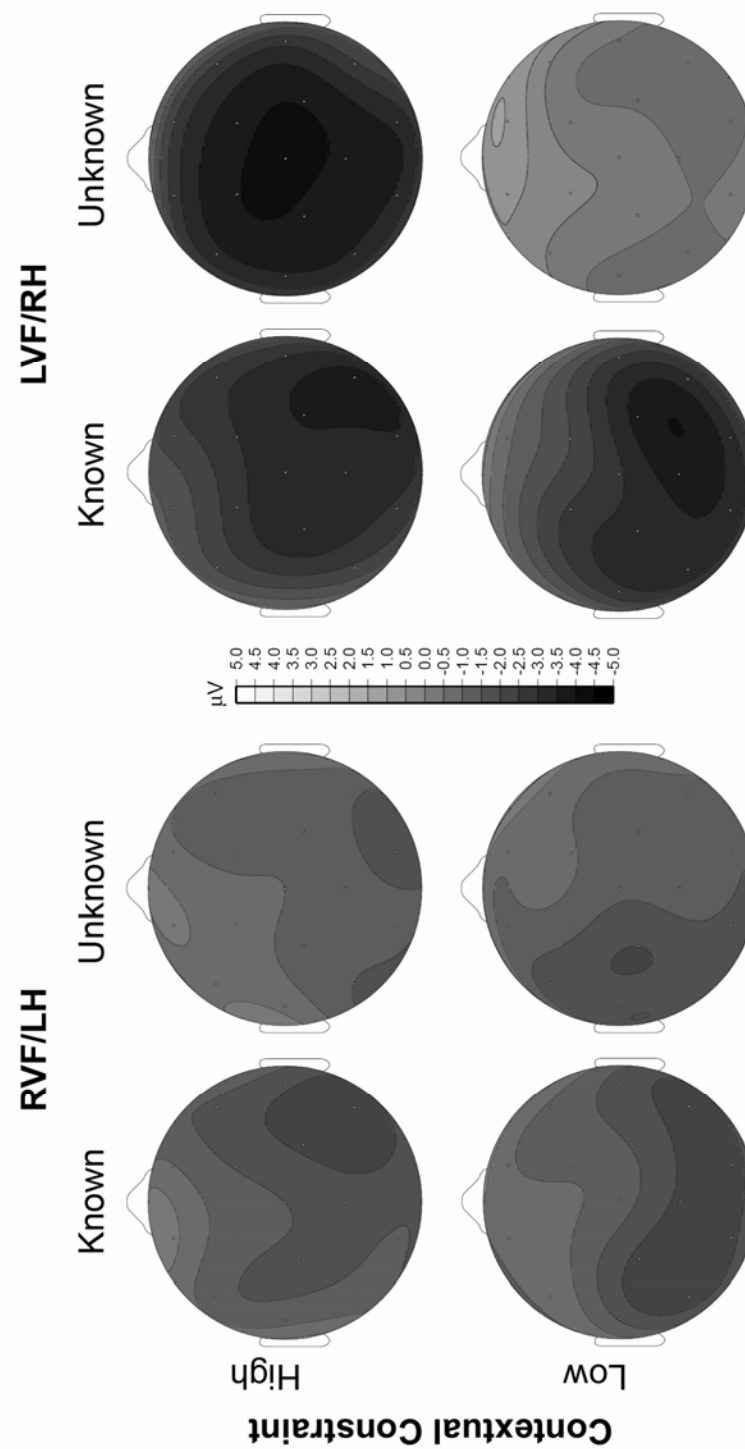


Figure 5-11. Spline interpolated isovoltage maps plot the distribution of the priming effect (Unrelated – Synonym/ID) from 500-900ms across experimental conditions in both visual fields.

5.4.3 Correlation between reaction time and N400

Since we found that N400 priming effects were not coincident with reaction time measure of priming in the Unknown word conditions, we decided to further explore the relationship between N400 amplitude and RT in the priming conditions. We examined the association between N400 and reaction time across all conditions by measuring the correlation between average N400 amplitude from 250-500ms at each electrode site to targets presented in all priming conditions for each visual field, LVF and RVF: Known/High/ID, Known/High/Unrelated, Known/Low/ID, Known/Low/Unrelated, Unknown/High/ID, Unknown/High/Unrelated, Unknown/Low/ID, Unknown/Low/Unrelated, with that of reaction time for each priming condition. Figure 5-12a plots the mean amplitude for N400 across average reaction time at a single vertex electrode for conditions in the LVF/RH and RVF/LH, while displays Figure 5-12b the R^2 correlation coefficients for this relationship across all electrodes on the scalp for targets presented to each visual field. As can be seen from these figures, N400 amplitude and reaction time were strongly negatively associated over a broad region on the scalp in both visual fields, such that faster reaction times were associated with more positive N400 amplitudes. In addition, the distribution of this association tended to be weaker at frontal electrodes in both visual fields.

5.4.4 Correlation between LPC and reaction time

Although our reaction time measures did not indicate any evidence of priming for Unknown words, we found significant LPC priming effects for all Unknown word conditions,

except Unknown/Low in the LVF/RH. In order to further examine the relationship between the LPC

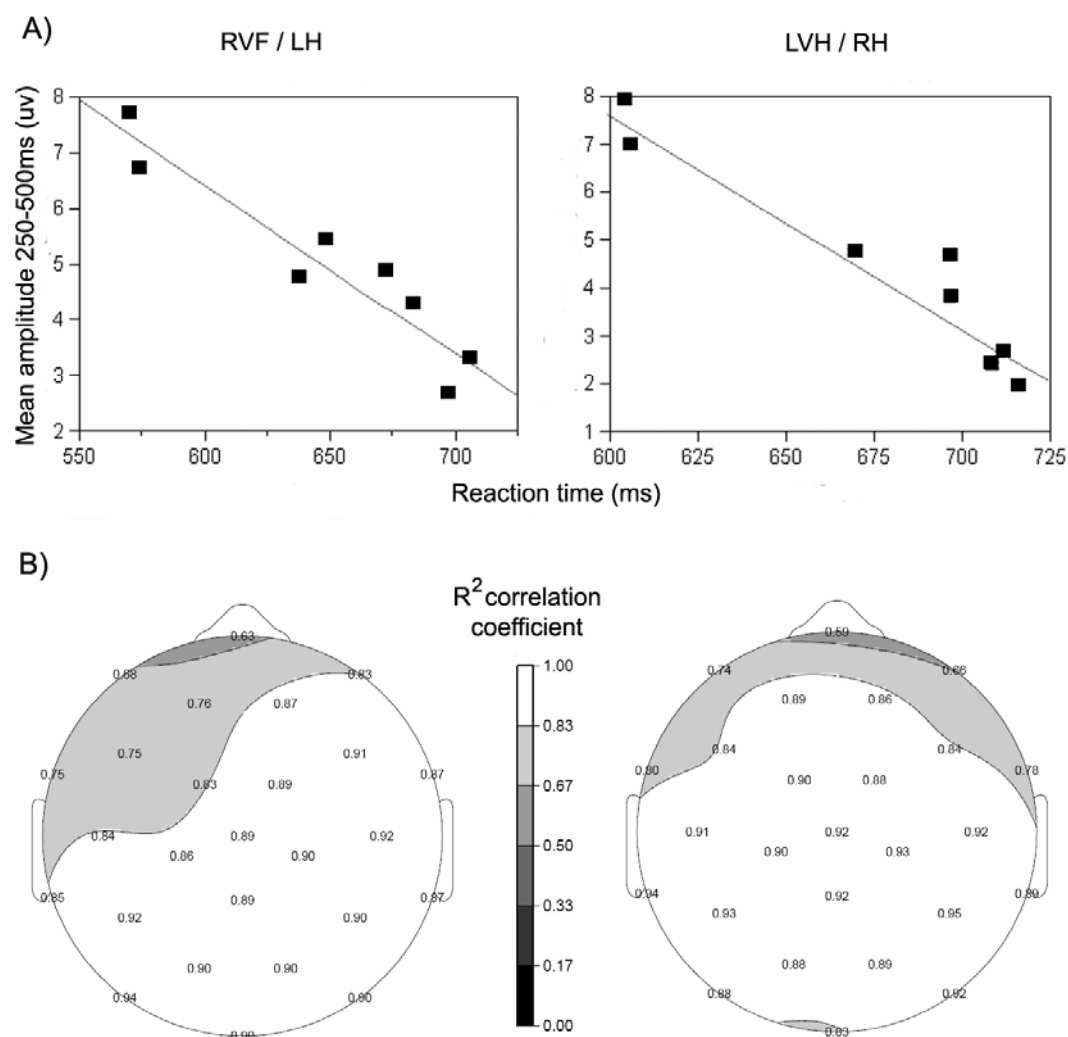


Figure 5-12. Correlation of reaction time and N400 across all priming conditions in each hemisphere in A) at a single vertex electrode (MiCe) and B) across all scalp electrodes.

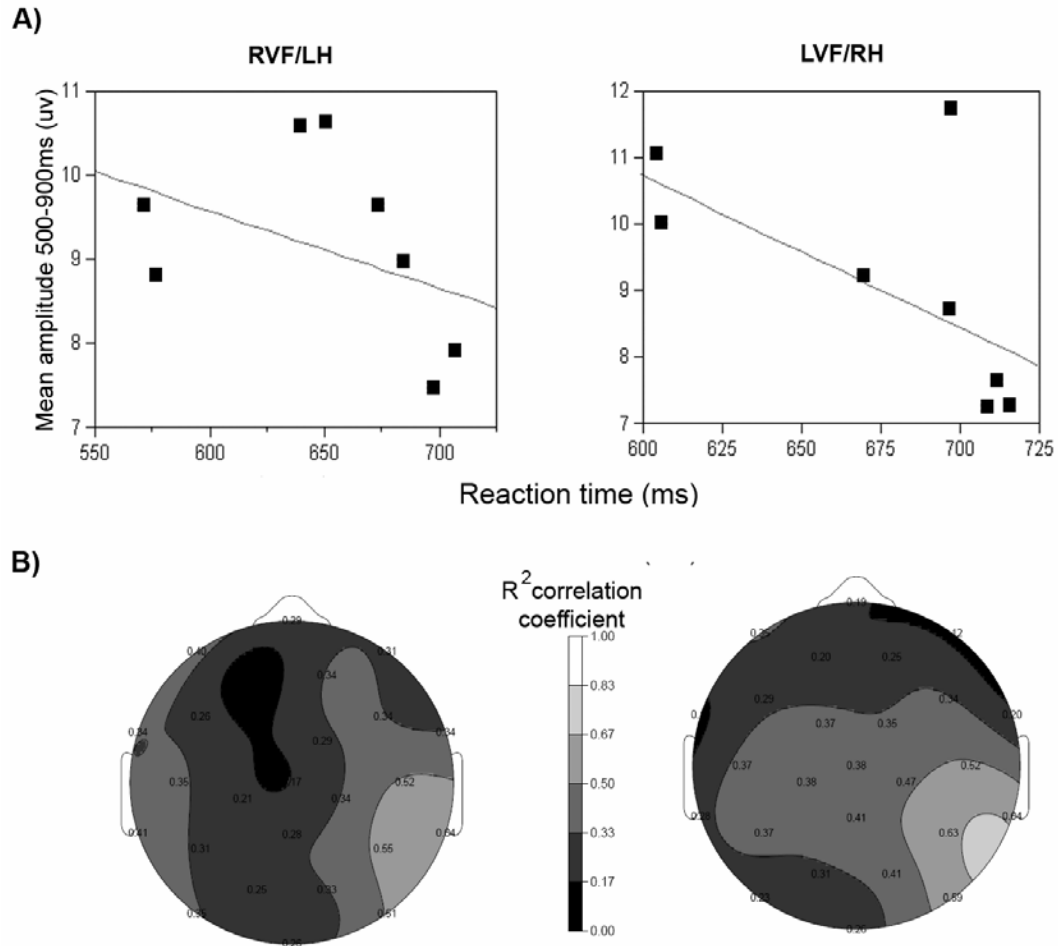


Figure 5-13. Correlation of reaction time and LPC amplitude across all priming conditions in each hemisphere in A) at a single vertex electrode (MiCe) and B) across all scalp electrodes.

priming effects and reaction time, we decided to examine the correlation between the LPC amplitude and RT in all priming conditions. We examined the association between LPC and reaction time by measuring the correlation between average LPC amplitude from 500-900ms at each electrode site to targets presented in all priming conditions for each visual field, LVF and RVF: Known/High/ID, Known/High/Unrelated, Known/Low/ID, Known/Low/Unrelated, Unknown/High/ID, Unknown/High/Unrelated, Unknown/Low/ID, Unknown/Low/Unrelated,

with that of reaction time for each priming condition. Figure 5-13a plots the mean amplitude for LPC across average reaction time at a single vertex electrode for conditions in the LVF/RH and RVF/LH, while displays Figure 5-13b the R^2 correlation coefficients for this relationship across all electrodes on the scalp for targets presented to each visual field. As can be seen from these figures, LPC amplitude and reaction time were negatively associated over a fairly focused region over right posterior electrodes. Here, faster reaction times were associated with more positive LPC amplitudes.

5.4.5 Nonword priming

Since a number of differences in priming effects were seen between Unknown and Known words, we decided to further explore if the response to nonwords varied when they were preceded by Known or Unknown words. In Figure 5-14 we compare nonword targets preceded by Known and Unknown words in all experimental conditions. From this figure, it appears that nonwords tend to show more positive amplitudes when preceded by Unknown words compared to Known words from approximately 200ms post stimulus onset to the end of the epoch. In order to examine how these waveform differences in nonword priming vary according to condition we measured N400 amplitudes between 250-500ms over all 26 channels, and LPC amplitudes between 500-900ms overall all 26 channels. For each component time window we conducted a 4-factor repeated measures ANOVA with factors of Word Type (Known or Unknown) x Constraint (High or Low) x Visual Field (RVF/LH, LVF/RH) x Electrode (26 level of electrode). For N400 amplitude, we found a main effect of Word Type [$F(1,23)=117.66$, $p<0.0001$], with nonwords preceded by Unknown words yielding more positive N400 amplitudes. There was also a main effect of Electrode [$F(25, 575)=42.94$,

$p < 0.0001$]. Even though a main effect of Word Type was observed, the only interaction of Word type with another factor was Word Type x Electrode [$F(25, 575) = 2.81$, $p < 0.0001$]. This was driven by a tendency for Word Type N400 differences to be largest at Centroparietal regions on the scalp. No other interactions were observed, indicating that the difference observed between N400 amplitudes for Nonwords preceded by Unknown and Known words did not vary by Visual Field, Constraint or Distribution, or any combination of these factors.

For LPC amplitude, we observed main effects of Word Type [$F(1,23) = 257.43$ $p < 0.0001$], Constraint [$F(1,23) = 5.48$ $p = 0.0193$] and Electrode [$F(25, 575) = 118.64$ $p < 0.0001$]. Main effects were driven by more positive LPC amplitudes for nonwords preceded by Unknown primes and High constraint primes, as well as a tendency for the LPC to be smaller at frontal electrode sites. An interaction for Word Type x Constraint was also observed [$F(1,23) =$], driven by larger (more positive) LPC amplitudes for nonwords preceded by Unknown words that appeared in both High and Low constraint contexts, compared to Known words, but for larger LPC amplitudes for Unknown/High, compared to Unknown/Low. Interactions of Word Type x Constraint x Visual Field and Word Type x Constraint x Visual Field x Electrode were not significant. This indicates that the difference observed in LPC amplitudes for Nonwords preceded by Unknown and Known words did not vary by Visual Field, Constraint or Distribution, or any combination of these factors.

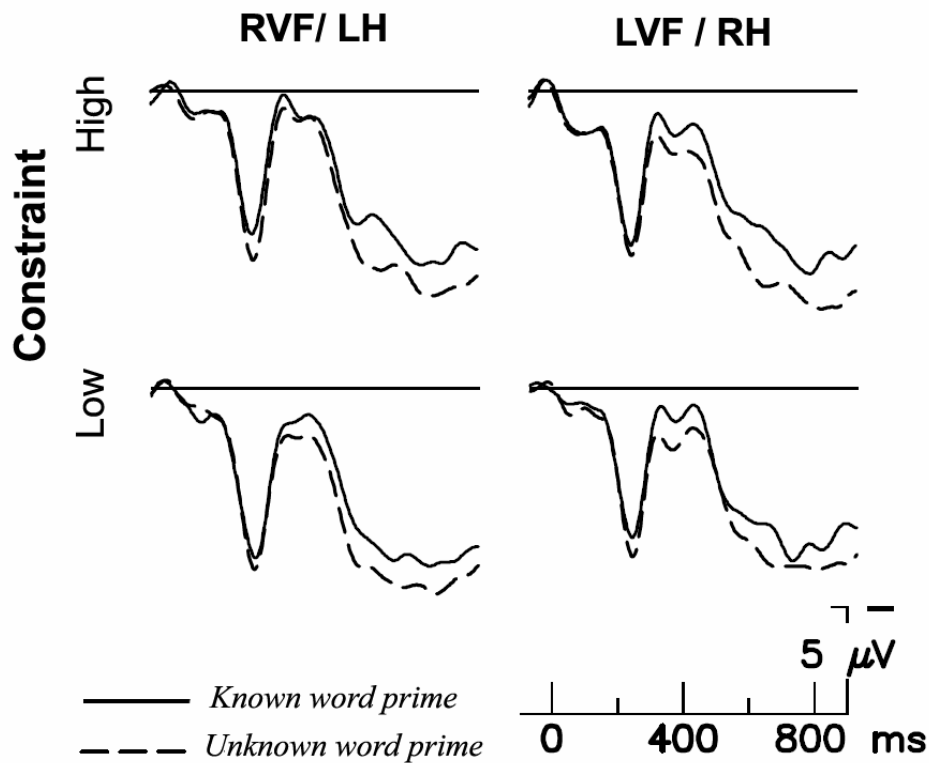


Figure 5-14. ERP waveforms to Nonwords preceded by Known and Unknown word primes at a single vertex electrode (MiCE).

5.5 Discussion

Previous work has found that a single presentation of a novel word in a highly constraining sentential context can result in a significant understanding of its meaning and usage, and that this understanding is reflected in electrophysiological measures of the N400 brainwave component. The aim of this study was to examine how this rapidly acquired representation of word meaning may become established across the cerebral hemispheres.

Our behavioral results replicated a number of standard findings. First, faster reaction times and increased accuracies were observed for targets presented to the RVF/LH

compared to the LVF/RH. This effect is a well documented outcome of the lateralized lexical priming task, and suggests a left hemisphere advantage. As expected, we also found standard priming effects for known words across both hemispheres. Known word primes that were identical in meaning to the target items (i.e. repetitions) yielded faster and more accurate responses to targets in both hemispheres compared to those that were unrelated in meaning. This is consistent with repetition priming effects in the left and right hemispheres (Marsolek, Kosslyn, & Squire, 1992; Weems & Zaidel, 2005).

By contrast, robust behavioral priming effects were not observed for Unknown (presumably recently learned) word primes. Reaction time comparisons failed to reveal priming effects for Unknown words in any condition. However, accuracy analyses revealed that targets preceded by Unknown word primes in the RVF/LH yielded a higher percentage of correct responses when the target item was a synonym. This effect was not observed via reaction time measures. These behavioral findings are also in accordance with previous word learning work that also failed to find behavioral priming effects to newly learned words (Borovsky et. al., in prep).

Our electrophysiological findings also replicated a number of standard results. For one, the visual N1 component was larger over electrodes contralateral to the visual field of stimulus presentation. This indicates that our paradigm was successful in directing the stimuli (initially) to a single hemisphere. Additionally, we found a number of typical N400 priming effects for stimuli presented to both hemispheres. N400 amplitudes were reduced when targets were preceded by Known words that were identical in meaning.

Although our N400 results were generally consistent with our behavioral results for Known words, they were significantly different for Unknown words. We found robust ERP

priming effects for targets preceded by Known words in both hemispheres. Differences due to prime relationship extended between 250-900ms, encompassing both the N400 and LPC time windows. In both time windows and hemispheres we observed typical priming effects. Less negative amplitudes were measured for target words that were identical or related to primes. In addition, the scalp distribution of these priming effects did not vary across experimental conditions.

By contrast, in the N400 time window, Unknown word priming was observed only in a single condition: namely, when target words initially presented to the left visual field/right hemisphere were preceded by Unknown words that had been learned in highly constraining contexts. These right lateralized N400 effects resembled findings using a non-lateralized (central field presentation) paradigm (Borovsky, Elman, & Kutas, in prep). With central presentation, we observed significant priming for Unknown words that were learned in highly constraining but not in less constraining contexts. Our findings suggest that the N400 effects observed with central presentation may have been primarily driven by representations encoded by the right hemisphere. Moreover, there were no reliable differences between the scalp distribution of Unknown/High priming effect in the LVF/RH and that of other prime conditions. This suggests that the neural correlates underlying the representation of novel word representations in the right hemisphere could potentially be similar to that of known words.

We also found a number of LPC priming effects to targets preceded by both Known and Unknown words. LPC effects in the Known word conditions were present across both constraint conditions and hemispheres. Targets identical in meaning to their preceding Known word primes elicited larger positivities than those that were unrelated in meaning.

Enhanced LPC amplitudes to synonymous targets were also observed in Unknown/High conditions in both hemispheres, as well as for Unknown/Low primes in the left hemisphere. This pattern differs from that in previous work in which Known and Unknown primes and targets were presented centrally. During central presentation extensive LPC effects such as the ones we observed here were not seen (Borovsky, Elman and Kutas, in prep). Extended positivities have been reported in a number of studies of sentential processing (Federmeier et. al. 2006, Coulson and Wu, 2006; Morena et. al 2002). In these cases, large late positivities have been elicited (especially over frontal electrodes) to items that were unexpected or unrelated to highly constraining sentences. These findings have been interpreted to suggest that the LPC indexes an increased allocation of cognitive resources after a violation of semantic expectancy (Federmeier et. al., 2006). Other late positivities have also been associated with violation of expectations of patterned and syntactic forms. It is unclear if these P3 and P600 effects are related to LPC effects, or even to each other (see Coulson, King and Kutas, 1998 for a review). However, these patterns suggest that late positivities tend to arise under conditions of expectation violation, surprise, or anomaly. Yet, unlike these previous findings, we observed increased positivities for targets that were identical in meaning to their Known and Unknown word primes – and not to unrelated targets. This is in accordance with word repetition studies which find increased LPC amplitudes with repeated words in sentences (Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991). In addition, priming studies have observed increased LPC amplitudes for related targets across both hemispheres (Bouaffre & Faida-Ainseba, 2007; Coulson, Federmeier, Van Petten, & Kutas, 2005). These studies do not support the idea that increased LPC amplitudes reflect violation of expectancy or surprise. Rather, Bouffre and Ainseba (2007) have suggested that this

pattern may reflect later conscious detection of the prime-target relationship. Given that our LPC effects were much stronger in Known than Unknown prime conditions, it is likely that participants may have detected the relationship between repeated Known words in the task. On the other hand, given that synonyms to Unknown words were not repeated, it is possible that participants did not as easily detect the semantic relationship between the two. Further research is necessary, however, to examine why these effects appeared in the lateralized and not centralized version of this task. One possibility is that since the LPC has been thought to index a number of cognitive processes, including semantic reintegration and comparison, it is also possible that the increased demands of the lateralized task were more difficult than the centralized version and this LPC effect may thus reflect extended comparison or processing of the target. Another difference between the centralized and lateralized version of this study is that the latency of the ERP effects appear later in the lateralized version. It is possible LPC effects may occur earlier in the centralized version, but overlap with the N400 component, thereby masking separable LPC effects.

Given that N400 priming effects were observed only in the right hemisphere for Unknown(newly learned)/High words, and not for any other unknown (newly learned) prime condition, this suggests that rapidly acquired representation of novel word meanings may be initially integrated into semantic memory via the right hemisphere. Our results are consistent with previous findings from the developmental literature, which have observed increased deficits in vocabulary acquisition in children who have obtained early injury to the right hemisphere (Eisele & Aram, 1993; Thal et al., 1991), as well as ERP studies that find greater bilateral activity at early stages of language learning (Mills, Coffey-Corina, & Neville, 1993). This suggests that the neural mechanisms involved in early word learning in children

may also extend to adults – since our findings suggest that the right hemisphere was primarily involved in priming of newly learned word meanings.

In the adult literature, word meaning representations in the left hemisphere have been characterized as “focused”, “fine” or “localized”, compared with “coarse”, “diffuse”, or “distributed” representations in the right hemisphere (Beeman, Friedman, Grafman, & Perez, 1994; Chiarello, 1998; Grose-Fifer & Deacon, 2004). In our study, it is possible that the initial representations that are established after a single presentation of a word in context may not have yet had the opportunity to develop a specific, fine-grained representation that is typical for the left hemisphere. Conceivably, such representations might require additional experience with a new word in many different contexts. This suggests that the right hemisphere may be very important in representing word meanings initially from context in single trial learning, but that the left hemisphere is important at later stages that involve continued experience with the word. This idea merits further study.

Neuropsychological studies of narrative and discourse processing in patient populations have also highlighted the importance of the right hemisphere. In this area, a number of studies have found that patients with right hemisphere brain injuries have relative difficulty in connecting extended narrative passages or discourse context, and that these deficits potentially stem from problems in inferring appropriate meanings from these contexts (Beeman, 1993; Tompkins, Scharp, Meigh, & Fassbinder, 2008). In our study, participants needed to rapidly integrate information from sentence contexts to infer appropriate meanings of novel words. Thus, it is possible that the involvement of the right hemisphere in priming of unknown word meanings could be related to the RH involvement of inferring appropriate meaning from context, but further work is necessary to test this idea.

In summary, this study is the first to examine the differential involvement of the cerebral hemispheres during acquisition of word meaning. Our electrophysiological results show a greater involvement of the right hemisphere in priming from of novel word meanings, and that novel word priming effects only appear when the novel word initially appeared in a highly constraining context. This work adds to previous findings that the N400 component is sensitive to acquisition of novel word meaning after just a single exposure in a highly constraining context. Thus, while previous studies have observed electrophysiological evidence of novel word priming after a single instance, this study indicates that this priming effect is driven primarily by the right hemisphere. With this study we add to a growing body of evidence that significant knowledge is acquired about novel word meanings after just a single instance, and we observe for the first time that the mechanism responsible for this rapid neural change is principally driven by the right hemisphere.

Chapter 5 appears in preparation for publication by Borovsky, Arielle, Elman, Jeff and Kutas, Marta. The dissertation author was the primary investigator and author of this paper.

CHAPTER 6

GENERAL DISCUSSION

Humans learn many thousands of words, often without explicitly trying, and after very limited experience with any individual word. Given the broad experience with language that every individual will gain over a lifetime, word learning must be at least partly influenced by linguistic experience. The nature of this experience can be many things, ranging from long-term experience to language over the lifespan, to immediate contexts realized via surrounding words in a single sentence. As reviewed in the introduction, there are a number of theories that attempt to explain the cognitive strategies that are used to constrain the potential meanings of novel words. These theories tend to examine how non-linguistic information or cognitive biases can influence linguistic learning. However, they often ignore the influence of the linguistic context itself on word learning. Accordingly, although word learning is a highly active research area for developmental and cognitive scientists, relatively little research has explored how linguistic experience can influence lexical acquisition. In this dissertation, I explore how two types of linguistic experience influence the cognitive and neural bases of word learning. The first type, examined in chapter 2, is the influence of lifetime language input on lexical acquisition. The second, examined in chapters 3, 4, and 5, is that of immediate sentential

context. Before turning to a more general discussion of this work, I first review and discuss the findings from each in the sections below:

6.1 Lifetime language experience

There is considerable evidence that the structure and quantity of a child's early language experiences can shape vocabulary acquisition. One of the earliest and most influential investigations of the association between language input and vocabulary learning was conducted by Betty Hart and Todd Risley (1995). In their work, they found a direct and striking correlation between the amount of speech that children hear at home and their eventual vocabulary outcomes. While the finding that increased language exposure in children leads to higher vocabularies is in itself unsurprising, the research is nonetheless immensely important because it documents the wide-ranging variety in language input heard by children. This suggests that children do not hear equivalent language input – and this has meaningful consequences for later language development. As reviewed in the introduction, there is substantial consensus that increased language exposure results in favorable vocabulary outcomes. However, it is not well understood why this occurs, and how structural aspects of linguistic input might also influence lexical learning.

In this dissertation, we proposed that one reason that language input may influence vocabulary acquisition, is that the input may shape semantic organization of words in the lexicon, and thereby influence subsequent language acquisition. This idea had been initially proposed by Gopnik and Meltzoff (1987, 1993, 1997), when a number of experiments found that the vocabulary spurt was associated with the ability to sort objects

into categories. However, these experiments only indirectly examine this hypothesis. There are a number of reasons why direct examination of this hypothesis is difficult to test with children. First, experiments with children can not easily probe the semantic organization of the lexicon, as the main method used to examine semantic organization utilizes speeded reaction time in adults. In addition collecting in-depth conversational data of child-directed speech is both expensive and time intensive, and thus, it is difficult to collect enough data to examine a wide variety of participants across many language backgrounds. Also, it is not possible to substantially alter the language exposure of a child on a large scale. We therefore chose to examine this hypothesis directly by using computational simulations.

In a series of simulations, we examined how three properties of language input might influence later lexical learning and development of semantic category structure. These properties – amount, frequency, and structure – account for some important and salient types of variation that might be plausibly seen across children. Below, I summarize the results from all three studies:

6.1.1 Amount of input

In this simulation, we manipulated the overall variety of input on which networks were trained. This input ranged in corpus size from 20 to 1000 sentences. In general, networks trained on larger corpora showed improved ability to acquire novel words, and improved category development. In fact, there was a highly significant correlation between the category structure measurements and speed of lexical acquisition. Together, these results are in line with previous findings in children, which find that increased

experience with language results in larger vocabularies. However, this research also sheds light on the potential reasons why increasing language exposure also improves word learning in children. Rather than observing a pattern where greater experience hones language acquisition ability, we see that the input influences the cognitive structure of the lexicon. Increasing input allows the relationship between words that are similar in meaning and use to be realized earlier, thereby improving the subsequent learning of other words that also belong in the same category.

One potential criticism of this experiment is that the range of sentences used (20 to 1000) and vocabulary size (95 words) does not represent a realistic variety in the words that children encounter. However, the intended purpose of this simulation was to examine the relationship between linguistic input and semantic structure. This question does not require a full-blown natural language, and is more easily understood when using this restricted and well controlled artificial language. However, in order to confirm the findings of this study, there are a number of questions that could be addressed in future work. A first and obvious experiment would be to examine the relationship between appearance of sorting abilities and amount of speech heard by children in a series of typical play sessions with their parents. These play session speech measurements could replicate the extensive home visits of Hart and Risley (1995), and provide a representative window into the overall language input heard by these children. If there is relationship between onset of sorting abilities and language exposure is found, then it would be expected that children who are exposed to more language input would show an earlier onset in their ability to sort objects into categories. If this is found, it would provide a useful replication in children and confirmation of the relationship observed in our networks between semantic category

structure and language input. Additionally, the advancement of electrophysiological recording techniques in children has also made it possible to examine infant's understanding of semantic category relationships between words. Recently, Torkildsen and colleagues (2006) have explored the relationship between vocabulary size and semantic organization in 20 month old children by using electrophysiological methods. They find that children with high and low productive vocabularies show some small differences in ERP responses to semantic picture/word mismatches. However, these differences were not statistically significant. Part of the reason for the insignificant findings between vocabulary groups was that children did not produce all of the words used in this study. On average, each child had 27 out of 90 total words used in the experiment in their productive vocabulary. It is possible that the additional noise added by the general unfamiliarity with the words that appeared in the study could have masked potential vocabulary group differences in semantic organization. Despite these methodological problems, the experiment suggests a method by which the relationship between semantic organization, vocabulary level, and language input could be explored in future work with children.

In sum, this simulation reveals a significant relationship between semantic development and total amount of input to which a neural network is exposed. Together this suggests that there is potentially important relationship between early language exposure and cognitive development that can influence subsequent language ability. However, in order to confirm this relationship, it is necessary to explore this question in experiments with children. Advances in ERP and other real-time language processing measures that do not require overt responses from children are providing exciting

advances in the field and can provide additional information regarding the relationship between language input and cognition.

6.1.2 Frequency of input

This simulation compared input that varied in the frequency distribution of words within categories. In one condition, all words were presented with roughly equal frequency. In the other, each category contained a single, highly frequent exemplar, while the rest of the words in the category were presented with relatively low frequency. After training, we observed that the networks more readily learned novel words, and more quickly developed highly organized categories. The networks exhibited improved learning rates in the uneven category because they were able to generalize information learned about the single highly frequent exemplar more quickly to other category members. There is some indication that children will also learn more effectively under these conditions of uneven frequency distribution. For instance, studies by Goldberg and colleagues have found that verb constructions are learned more easily when trained with a highly frequent exemplar. However, as of yet, no work has addressed how the frequency distribution of words within a semantic noun category might influence the acquisition of words in that category. One indication that this may also be helpful comes from research by Snow and colleagues (1998) that find that the presence of infrequent, or rare words in parental speech aids in child vocabulary learning. It could be possible that the inclusion of rare words serves a similar function in the uneven frequency condition. In the case of the simulations, frequent and infrequent words were incorporated into the semantic structure of the network earlier than the network where all words appeared with equal frequency due to the fact that a single

highly frequent word was acquired especially early. Perhaps providing numerous infrequent words in speech to children also capitalizes on this boost in learning. In this case, our simulation suggests that categories in which the child already understands at least a single member quite well should show facilitated acquisition of novel, infrequent words compared to categories in which a number of words are only partially understood. Future work is necessary to test this hypothesis with children.

6.1.3 Structure of input

In this study we manipulated the grammatical complexity of linguistic input by comparing two conditions, one complex and one simple. In the simple condition, networks were exposed to input that contained simple transitive (NVN) and intransitive sentences (NV). Complex sentences contained simple plus two additional complex constructions – ditransitive (NVNN) and matrix (NVNVN). After training, the simple input networks were faster at learning words, and showed improved semantic development. Once again, we found that the relationship between semantic development, lexical acquisition, and linguistic input was tightly related.

These results augment previous work that has examined the relationship between the structural complexity in child directed speech and lexical development. These previous studies have reported conflicting findings, with some suggesting that simpler input is beneficial for lexical acquisition (Brent & Siskind, 2001), and others suggesting it is not (Hoff & Naigles, 2002). In our case, we find the simpler input is beneficial – at least at the earliest points in development – and that the reason it is beneficial is that it allows for networks to more easily appreciate the relationships between words that are similar in meaning.

Essentially, when words are utilized in a reduced number of grammatical constructions as they were in our “simple” input condition, networks identified the similarities between words in the same semantic category more efficiently than in the “complex” condition with many constructions.

Even though simpler input was beneficial at the early stages in language development simulated by this model, it is possible (and perhaps likely) that the inclusion of additional structural complexity may pose numerous benefits to a child at later points in development. This may in part explain why conflicting results have been reported between observational studies of CDS complexity. It is necessary that future work explore this question empirically, rather than observationally, by training novel words in sentence constructions that are simple and complex. An additional hypothesis suggested by our work predicts that having highly developed knowledge about members within a category will provide an additional benefit when learning words from simple and complex constructions. In fact, it may be possible that children may only successfully learn words in complex constructions when they have highly developed knowledge of words in the same category. This idea merits further research.

6.1.4 Conclusions

The results of these simulations reveal three main findings: 1) the ability to acquire new words can be influenced by previous experience in the structure, frequency and amount of linguistic input, 2) semantic development is associated with word learning ability and 3) there is a tightly associated relationship between semantic development and linguistic experience. In sum, our findings reveal that variation in linguistic input between

children has dramatic consequences in the cognitive arena of word learning. Children who have been exposed to input that improves word learning are not better learners just because they have more or “better” experience – its because the underlying cognitive mechanisms involved in word learning have changed as a result of that input. In addition, we have demonstrated that these changes in cognitive development can be characterized via a learning mechanism that remains the same across development. In this case words are learned by the way in which they are used such that words that are used in similar sentence contexts come to be grouped into categories according to similarity in meaning and usage. This kind of learning mechanism could plausibly account for word learning in children, and it does not specify the need for special cognitive machinery, biases, or introduction of novel learning mechanisms at the onset of the vocabulary spurt. Rather, this mechanism suggests that improvements in word learning ability occur due to development in the child’s understanding of the relationship between word meaning and usage. Specifically, when a child understands the relationships between words in the same category, subsequent learning in that category is improved. This mechanism suggests that word learning does not improve uniformly, and that it may be possible to observe improved learning in some categories which children understand very well, compared to those that are less well known.

In addition, the linguistic input was critically important in the development of these networks. Networks that were exposed to a greater variety of input, simpler sentences, and highly frequent exemplars exhibited improved learning of both category structure and novel words. This extends previous findings that suggest that increased input is important to child word learning. Rather than finding that only one aspect of the input matters, we

find that there are at least several (and potentially more) factors in language input that can influence word learning ability. This is encouraging because it suggests several potential “shortcuts” for language development for children who have not had the benefit of exposure to an enriched language environment early in life. Since it is difficult for children to catch up in deficits in total language exposure, it is instead possible to present other advantages in the reduced input the child does here by speaking in sentences that are simpler in structure, and is peppered with a few infrequent words, but many highly frequent words.

6.2 Immediate sentence context

The meaning of a word must be shaped by the context in which it appears. Although the relevant “context” can take many forms, the majority of any individual’s vocabulary will be learned implicitly via linguistic contexts during reading and speaking (Jenkins, Stein, & Wysocki, 1984; Nagy, Anderson, & Herman, 1987; Nagy, Herman, & Anderson, 1985; Sternberg, 1987). It is therefore important to understand how language contexts can influence the representation of novel words. Since word learning can occur very rapidly, a number of questions arise regarding how the brain rapidly forms these representations, and what information is represented about word meaning. There is a growing body of research to suggest that the electrical brain activity measured at the scalp is sensitive to changes in neural activity due to word learning. The first of these studies, reported by McLaughlin, Osterhout and Kim (2005) explored second language learning in adults learning a second language in a college French course. They found that over the course of several months their brains came to recognize the difference between words that were and were not actual

French words, as well as the difference between two words in French that were related in meaning (*chien-chat* “dog-cat”) or unrelated in meaning (*chien-nez*, “dog-nose”). Moreover, the N400 component of these student’s brainwave responses during the lexical decision task were identical to that of native French speakers’ by the end of the course. Interestingly, behavioral performance on the task did not reveal increased sensitivity to distinguishing real from nonwords during the lexical decision responses as students increased their experience with the language. This research is important because it suggests that the brainwave activity is sensitive to acquisition of word meaning even when the learner is not consciously aware of the improvement. Moreover, since the majority of word learning occurs without explicit training, but rather incidentally in context, it suggests that electrophysiological methods may provide a more sensitive measure of lexical acquisition than many behavioral methods.

In the experiments described in Chapters 3, 4, and 5, we measure brainwave activity as participants view novel words in sentence contexts that are either strongly or weakly constraining. Then, participant’s knowledge of these novel words are probed in one of two tasks: 1) a plausibility judgment task where novel words reappear as the objects of transitive verbs or 2) a lexical decision task, where novel words reappear as primes that are synonymous, related, or unrelated in meaning to known target words. The first task is designed to probe how a single presentation of a word in context influences the acquisition of word usage. In order to successfully discern the difference between appropriate and inappropriate word usage in this task, participants must have integrated significant and relatively sophisticated information about the novel word’s relationship with other verbs. In the second task, knowledge of word meanings are probed in a less explicit manner. Rather than asking participants to make decisions about novel word usage, participant’s

understanding of novel word meanings is probed via a semantic priming task using a lexical decision. In this task, participants never make decisions about the novel words themselves, or the relationship between the prime and target words. Instead, they are asked to make lexical decisions on known words that vary in semantic relatedness to the intended meaning of the novel words, and previously viewed known words. In this way it is possible to determine if participants understand the relationship between the meanings of the novel words and other related word meanings. Altogether, the studies in Chapters 3 and 4 investigate the influence of contextual constraint on the acquisition of word meaning and usage – both of which are important aspects of word learning. Chapter 5 extends this research by asking how the cerebral hemispheres participate in the acquisition of novel word meaning. In this study, participants once again viewed novel words in strongly and weakly constraining sentence contexts. Novel word meanings were then probed via a lateralized semantic priming task. Unlike in Chapter 4, where all words were presented to the central visual field, target words were presented to either the left or right visual field while participants performed a lexical decision. Using this task, we then compared brainwave responses to target words that appeared in both hemispheres in order to determine how the preceding prime might have affected their processing. In sum, the three experiments in Chapters 3 through 5 examine how sentential context influences what is learned about word meanings from context, and where the brain initially integrates these meanings into the mental lexicon. Below, I summarize the main findings from each of the studies in the order that they were described.

6.2.1 Learning to use words

As described above, the goal of the study in Chapter 3 was to examine how contextual constraint influences later ability to detect appropriate and inappropriate usages of a novel word. Success on this type of task requires that participants understand the relationship between the novel word and potential verb meanings that could and could not plausibly be associated with the word.

We found a number of interesting findings. First, as expected, we found that the brain does distinguish between appropriate and inappropriate usages of already known words in this task. The time at which we saw this response however, might initially seem surprising. Rather than seeing the brainwave result at the time at which the known word was presented, this result appeared at the verb that preceded the known word. The reason that the N400 component differentiated between plausible and implausible usages of known words at a point preceding the actual presentation of the word, was that the form of the sentence was completely predictable, and the point of the verb was where all the necessary information to make a mental decision became available. This is because participants saw three sentences back to back that all ended in the same known or unknown word. The first of these sentences always provided the “context”, which the learner was asked to read and understand. In the second and third sentences, participants were then asked to decide if the novel word was being used appropriately. These sentences always followed a predictable form, of Subject– Verb – Target word. We therefore saw a difference between implausible and plausible usages of known words at the verb preceding the target word, when the target word was an already known word. This effect was equally large for known words that appeared in both High and Low constraint contexts, presumably because participants were

already highly familiar with the known words, and the sentence contexts did not provide significantly new or unexpected information about these word's meanings.

On the other hand, we did observe an effect of constraint for unknown words. Unknown words that initially appeared in highly constraining sentence contexts did show an N400 effect of plausibility, whereas words that appeared in low constraint sentences did not. This result suggests that participants were able to rapidly acquire enough information from just a single presentation of an unknown word in a high constraint context to detect the difference between appropriate and inappropriate usages of the word. In addition, this size of the N400 plausibility effect for these "high constraint" unknown words was identical to that for real words. This result differs markedly from our behavioral findings, where accuracy of the participant's button press decisions of sentence plausibility was reduced for novel words in both constraint conditions. Once again, like previous work, this result suggests that behavioral responses may not always be as sensitive a measure to acquisition of word meaning as that of brainwave responses.

6.2.2 Integrating words into semantic memory

The aim of the study in Chapter 4 was to extend the findings in Chapter 3 by further examining how contextual constraint influences the acquisition of word meaning. In this study, rather than measure understanding of appropriate word usage, we probed word knowledge in a less explicit manner. Novel words served as primes to known word targets that varied in relatedness to the intended meanings of the novel words. In this way, we were able to explore how novel word meanings become integrated into semantic memory when presented in contexts of varying contextual constraint.

The results both confirmed and extended our findings from chapter 3. As expected, known words that had appeared in contexts of both low and high constraint showed large reaction time and N400 differences when they primed target words that were identical and unrelated in meaning. Target words that were related in meaning also elicited N400 priming effects when compared to identical targets, but showed less reliable priming differences when compared to words that were unrelated in meaning. This effect is not unusual, and may in part reflect the fact that we chose related targets that were only categorically related (such as apple-GRAPe), but not associated in meaning (like mouse – CHEESE). Generally priming effects are weaker when associated stimuli are not included. Similarly, robust reaction time priming effects were observed when comparing identical stimuli to either related or unrelated targets, but related and unrelated targets did not show large differences in reaction time responses. Taking the behavioral and ERP results together for known words, this would suggest that the N400 reflects similar cognitive processes as that probed in the semantic priming lexical decision task.

However, our priming results for unknown primes suggest a different story. Unknown words that had initially appeared in highly constraining contexts induced clear N400 priming effects, whereas those that had appeared in low constraint sentences showed an absence of N400 priming effects. This is similar to our findings in the first study, where novel words in the high constraint condition elicited significant N400 plausibility effects, while these effects were non-existent in the low constraint condition. On the other hand, novel words in both constraint conditions did not elicit behavioral priming effects. Once again, this is similar to our findings in Chapter 3, which also suggested that a single presentation of the novel words was not sufficient to acquire significant information about

its usage, irrespective of the initial context in which it appeared. However, it is clear from this study that this knowledge extends to an understanding of semantic relationships between words.

6.2.3 Word learning across the hemispheres

Chapter 5 extended our work in Chapter 4 by examining how semantic relationships between novel words and known words become integrated in the cerebral hemispheres. This study was in part motivated by numerous findings which suggest that the different hemispheres represent and encode differential information about word meanings. Generally, representations in the left hemisphere have been defined as “focused,” “fine” or “specific”, whereas those in the right hemisphere have been characterized as “difficuse”, “coarse” or “broad”. In this study, we examined how the hemispheres differentially integrate novel words into semantic memory.

The task was very similar to the task in Chapter 4. Participants once again viewed known and unknown words in high and low constraint sentences. They then subsequently saw these known and unknown words as primes in a semantic priming task. Targets were either synonymous or unrelated in meaning to the primes, and were presented either to the left or right visual field.

As expected, we found N400 and behavioral priming effects for targets that were preceded by known word primes in both visual fields. Once again, priming was not observed for novel words that initially appeared in low constraint contexts. Interesting differences between the centralized and lateralized version were found for targets preceded by unknown words that had been previously seen in high constraint sentence contexts. In

Chapter 4, we found that these unknown words induced a modulation in the N400 when preceding identical and unrelated targets. In this study, we found that this N400 priming effect only appeared when targets were presented to the left visual field / right hemisphere. This finding suggests that the right hemisphere initially participates in the integration of novel word meanings into semantic memory. Since words which are already known show priming in both hemispheres in this task, this finding also suggests that it may be possible that additional experience with unknown words might be necessary for semantic priming to appear in both hemispheres for newly learned words. On the other hand, it is also possible that the left hemisphere might initially represent other kinds of information about novel word meaning that were not probed in this task. At the very least, these findings indicate that the hemispheres do not initially encode the same information about novel word meanings and that the right hemisphere seems to recognize the relationship between novel word meanings and other words earlier than the left hemisphere.

6.2.4 Conclusions

The results of the brainwaves studies reveal a number of important findings including: 1) contextual constraint influences acquisition of word meaning, 2) part of what is rapidly learned about word meanings includes both knowledge of how a word is used and its relationship to other word meanings and 3) the N400 brainwave component is sensitive to the acquisition of this knowledge. In sum, our findings suggest that the sentential context in which an unknown word appears has important consequences for its acquisition, and our experiments reveal a method by which this acquisition can be measured.

There are numerous advantages to using the ERP method to measure word meaning acquisition. For one, our studies and others have found cases in which standard behavioral measurements, do not reveal acquisition of word meaning, even when modulations in the N400 would indicate otherwise. Secondly, since ERP methods do not require a response from participants, it is possible to gauge acquisition of word knowledge in a broader set of populations. For instance, toddlers are unable to comply with tasks which require speeded button presses, however many interesting language development phenomena occur in children of this age. Even though our studies explore lexical acquisition in adults, it is possible that a modified version of our method could be extended to studies in children, or even in adult populations who are recovering language ability after brain injury.

Our studies seemed to reveal acquisition of word meaning even in cases where behavioral responses did not. Why were behavioral responses less sensitive in this case? It is perhaps possible that other behavioral paradigms might be discovered that could be just as sensitive. Our studies used some fairly standard psycholinguistic behavioral measurements of real time language processing. It is possible that these tasks might also utilize other cognitive mechanisms that are not measured by ERPs. In this case, it might be more appropriate to say that reaction time measures and ERPs measure highly overlapping, but not identical processes, and the processes measured by our behavioral tasks are not involved in rapid acquisition of word meaning.

There are a number of questions opened by this research towards which future work could proceed. We explored a fairly limited set of knowledge about word meaning – knowledge of usage and relationships to other words. It is possible that other aspects of word knowledge might take longer to acquire. For example, it is often possible to describe a

number of relevant features of nouns. Featural knowledge about a dog might include “is an animal,” “has four legs,” “barks,” and so on. Nouns can also be often described by a number of adjective features, such as by color and size. Future studies might explore these other aspects of word meaning acquisition from context in order to gain a more complete picture of how and when novel words become fully functional and integrated into our mental dictionaries. Additionally, our studies only examined how words are acquired after a single presentation. However, it is often the case that individuals will learn information about words gradually from repeated exposures. Other studies could explore how repeated exposures influence the acquisition of word knowledge either from repeated exposures in multiple unrelated contexts, or even in coherent discourse. Finally, we studied word learning in a relatively restricted population – that of college students. Yet word learning is a lifelong task, and it remains to be known how context might differentially influence word acquisition in other populations and interact with knowledge and experience.

In sum, in this dissertation we have identified a method to measure word learning as it naturally occurs in one’s native language. We have found that the contextual constraint in which a word appears can influence the knowledge that is acquired after just a single presentation of an unknown word. This knowledge includes both an understanding of the word’s appropriate usage and its relationship to other word meanings. In addition, we found that the cerebral hemispheres participate differently in this process. However, as reviewed above, there are still many unexplored questions that need to be answered before we will have gained a full understanding of the role of context in word acquisition. It is my hope that this method will continue to be utilized to reveal new and exciting facets of word meaning acquisition.

APPENDIX A

AVERAGE PRECISION

Average precision is calculated in several steps. First, the Euclidean distance between each words hidden unit vector and every other word is calculated. This is to find the pair-wise distance between each word and every other in representation space. The values are then ranked, so that for each word the other words have the most similar hidden unit representation are ranked closer than those that are not. Then, to calculate the average precision of each word is calculated with the following formula:

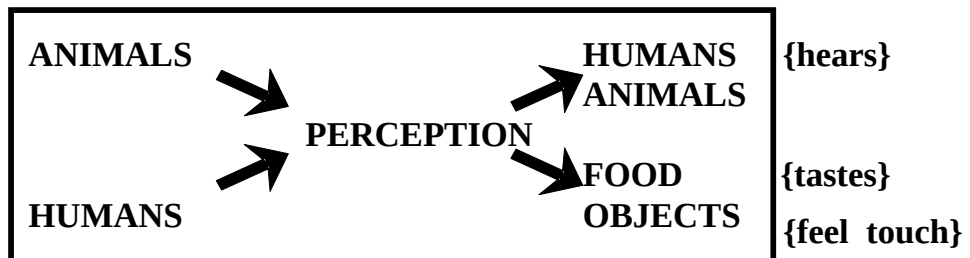
$$P(w) = \frac{1}{|C_w|} \sum_{i \in C_w} \frac{|n_{wi}(C_w)|}{|n_{wi}(all)|}$$

Where $P(w)$ stands for the average precision of one word, C_w is the number of words in the category, n_{wi} is the rank number of a particular word, and $n_{wi}(C)$ is the number of words in the target words category that have appeared before the particular rank. Simply, this algorithm calculates the proportion of words that belong to a target words category at each rank, and then divides this proportion by the number of words in the category. In a best case scenario, where all the words in a category are the closest ranked members, this would yield a value of one. Theoretically, these values can approach zero, where all within category words are infinitely far away from the target word. However, in practice, this result is difficult to achieve, so normally, just by chance, average precision values will hover around 0.2 without any real structure.

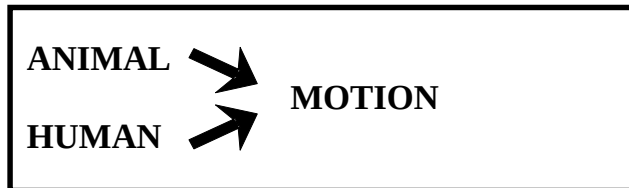
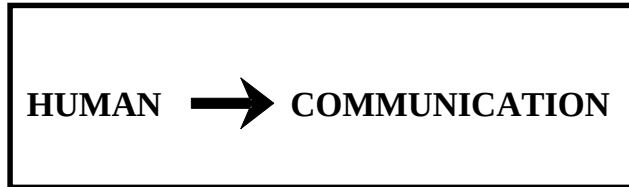
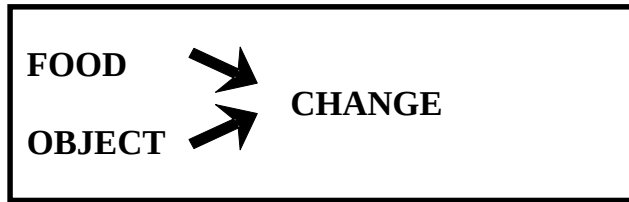
APPENDIX B

SEMANTIC RELATIONSHIPS BETWEEN CATEGORIES

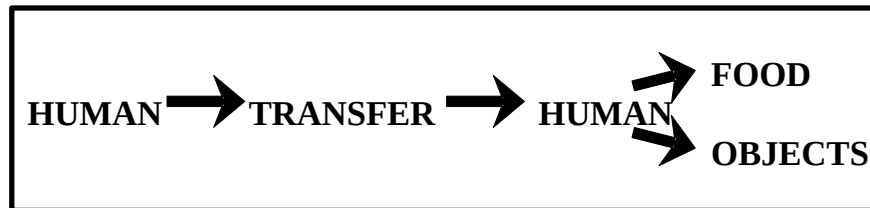
Transitive Relations



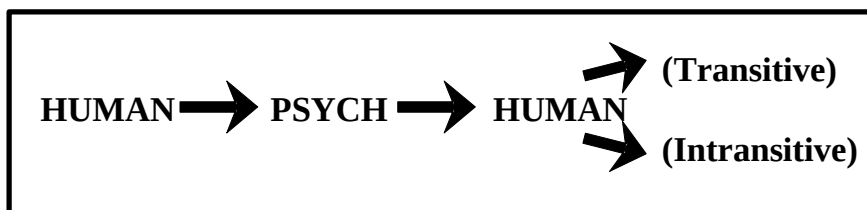
Intransitive Relations



Ditransitive Relations



Matrix Relations



APPENDIX C

EXAMPLES OF SENTENCES USED IN THE STUDY

Transitive Sentences

EATING:

kid gobbles pizza .

bird drinks water .

PERCEPTION:

rabbit sees bread.

grandmother touches book .

ACTION:

animal bites puppy .

teacher hugs lamb .

Intransitive Sentences

CHANGE:

cake falls .

box breaks .

COMMUNICATION:

boy talks .

kid laughs .

MOTION:

tiger moves .

mother jumps .

Ditransitive Sentences

TRANSFER

mother buys brother cup .

brother offers grandfather box .

Matrix Sentences

PSYCH

uncle wants grandmother sees duck .

grandfather convinces brother sits .

REFERENCES

- Aitchinson, J. (1994). *Words in the mind: An introduction to the mental lexicon* (2nd ed.). Oxford: Blackwell.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247-264.
- Anderson, J. E., & Holcomb, P. J. (1995). Auditory and Visual Semantic Priming Using Different Stimulus Onset Asynchronies - an Event-Related Brain Potential Study. *Psychophysiology*, 32(2), 177-190.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary Knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.
- Anglin, J. M. (1993). Vocabulary development: A morphological analysis. . *Monographs of the Society for Research in Child Development*, 58(10, Serial No. 238).
- Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Dev*, 62(5), 875-890.
- Bassano, D. (2000). Early Development of Nouns and Verbs in French: Exploring the Interface between Lexicon and Grammar. *Journal of Child Language*, v27 n3 p521-59 Oct 2000.
- Bates, E., & Carnevale, G. F. (1993). New directions in research on language development. Special Issue: Setting a path for the coming decade: Some goals and challenges. *Developmental Review*, 13(4), 436-470.
- Bates, E., & Goodman, J. C. (1997). On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition, Aphasia and Real-Time Processing. *Language and Cognitive Processes*, v12 n5-6 p507-84 Dec 1997.
- Bates, E. B., Inge, & Snyder, Lynn. (1988). *From first words to grammar: individual differences and dissociable mechanisms*. Cambridge: Cambridge University Press.
- Bauer, P. J., Dow, G. A., & Hertsgaard, L. A. (1995). Effects of prototypicality on categorization in 1- to 2-year-olds. . *Cognitive Development*, 10, 43-68.
- Beck, I. L., & McKeown, M. G. (1991). Conditions of vocabulary acquisition. In R. Barr, M. Kamil, P. Mosenthal & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 789-814). New York: Longman.

- Beeman, M. (1993). Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. *Brain & Language*, 44(1), 80-120.
- Beeman, M., & Chiarello, C. (Eds.). (1998). *Right hemisphere language comprehension: Perspectives from cognitive neuroscience*: Lawrence Erlbaum Associates, Inc., Publishers; Mahwah, NJ, US.
- Beeman, M., Friedman, R. B., Grafman, J., & Perez, E. (1994). Summation priming and coarse semantic coding in the right hemisphere. *Journal of Cognitive Neuroscience*, 6(1), 26-45.
- Behrend, D. A., Scofield, J., & Kleinknecht, E. E. (2001). Beyond fast mapping: young children's extensions of novel words and novel facts. *Dev Psychol*, 37(5), 698-705.
- Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain & Language*, 31(2), 308-327.
- Bentin, S., Kutas, M., & Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, 30(2), 161-169.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials associated with semantic priming. *Electroencephalography and Clinical Neurophysiology*, 60, 343-355.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials associated with semantic priming. *Electroencephalography and Clinical Neurophysiology*, 60, 343-355.
- Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: The M.I.T. Press.
- Bloom, P. (2000). *How Children Learn the Meanings of Words* Cambridge, MA: MIT Press.
- Bloom, P., & Markson, L. (2001). Are there principles that apply only to the acquisition of words? A reply to Waxman and Booth. *Cognition*, 78(1), 89-90.
- Borovsky, A., Elman, J., & Kutas, M. (2007). *Getting the gist is not enough: An ERP investigation of contextual word learning*. Paper presented at the Proceedings of the 29th Annual Cognitive Science Society, Nashville, TN.
- Borovsky, A., Elman, J., & Kutas, M. (in prep). Once is enough: N400 indexes semantic integration of novel word meanings form a single exposure in context.
- Bouaffre, S., & Faita-Ainseba, F. (2007). Hemispheric differences in the time-course of semantic priming processes: Evidence from event-related potentials (ERPs). *Brain and Cognition*, 63(2), 123-135.

- Bowerman, M. (1996). Learning how to structure space for language: A cross-linguistic perspective. In P. Bloom, M. A. Peterson, L. Nadel & M. F. Garrett (Eds.), *Language and Space* (pp. 385-436). Cambridge, MA: MIT Press.
- Braine, M. D. S. (1963). On Learning the Grammatical Order of Words. *Psychological Review*, 70(4), 323-348.
- Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.-F., Sommer, J., Wolbers, T., et al. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage*, 25(3), 958-968.
- Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2), B33-44.
- Broen, P. A. (1972). *The verbal environment of the language-learning child*. (Vol. 17). Washington, DC: American speech and hearing association.
- Brown, C., & Hagoort, P. (1993). The Processing Nature of the N400 - Evidence from Masked Priming. *Journal of Cognitive Neuroscience*, 5(1), 34-44.
- Brown, R. (1973). *A First Language: The early stages*. Cambridge, MA: Harvard University Press.
- Brown, R. W. (1958). *Words and things*. Glencoe, IL: The Free Press.
- Brownell, H. (1988). Appreciation of metaphoric and connotative word meaning by brain damaged patients. In C. Chiarello (Ed.), *Right hemisphere contributions to lexical semantics* (pp. 19-31). New York: Springer-Verlag.
- Brownell, H., Michel, D., Powelson, J., & Gardner, H. (1983). Surprise but not coherence: Sensitivity to verbal humor in right-hemisphere patients. *Brain & Language*, 18(1), 20-27.
- Burgess, C., & Simpson, G. B. (1988). Cerebral hemispheric mechanisms in the retrieval of ambiguous word meanings. *Brain & Language*, 33(1), 86-103.
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10, 425-455.
- Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual Differences in the Inference of Word Meanings From Context: The Influence of Reading Comprehension, Vocabulary Knowledge, and Memory Capacity. *Journal of Educational Psychology*, 96(4), 671-681.
- Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science: A Multidisciplinary Journal*, 27(6), 843 - 873.

- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17-29.
- Casasola, M. (2008). The Development of Infants' Spatial Categories. *Current Directions in Psychological Science*, 17(1), 21-25.
- Caselli, C., Casadio, P., & Bates, E. (1999). A Comparison of the Transition from First Words to Grammar in English and Italian. *Journal of Child Language*, v26 n1 p69-111 Feb 1999.
- Casenhiser, D., & Goldberg, A. E. (2005). Fast-mapping of phrasal form and meaning. *Developmental Science*, 8, 500-508.
- Chiarello, C. (1998). On codes of meaning and the meaning of codes: Semantic access and retrieval within and between hemispheres. In M. Beeman & C. Chiarello (Eds.), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience*. (pp. 141-160). Mahwah, NJ: Lawrence Erlbaum Associates.
- Chiarello, C., Burgess, C., Richards, L., & Pollock, A. (1990). Semantic and associative priming in the cerebral hemispheres: Some words do, some words don't . . . sometimes, some places. *Brain & Language*, 38(1), 75-104.
- Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, 38(6), 967-978.
- Childers, J. B., & Tomasello, M. (2003). Children extend both words and non-verbal actions to novel exemplars. *Developmental Science*, 6(2), 185-190.
- Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: the influence of language-specific lexicalization patterns. *Cognition*, 41(1-3), 83-121.
- Choi, S., McDonough, L., Bowerman, M., & Mandler, J. M. (1999). Early sensitivity to language-specific spatial terms in English and Korean. *Cognitive Development*, 14, 241-268.
- Chomsky, N. (1980). *Rules and representations*. London: B. Blackwell.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407-428.
- Cornelissen, K., Laine, M., Renvall, K., Saarinen, T., Martin, N., & Salmelin, R. (2004). Learning new names for new objects: Cortical effects as measured by magnetoencephalography. *Brain and Language*, 89(3), 617-622.
- Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, 50(4), 491-511.

- Coulson, S., Federmeier, K. D., Van Petten, C., & Kutas, M. (2005). Right hemisphere sensitivity to word and sentence level context: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31(1), 129-147.
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language & Cognitive Processes*, 13(1), 21-58.
- Courage, M. L., & Howe, M. L. (2002). From infant to child: the dynamics of cognitive change in the second year of life. *Psychol Bull*, 128(2), 250-277.
- Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Dev Psychol*, 33(6), 934-945.
- Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments & Computers*, 28(1), 125-127.
- Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, 92(1-2), 179-229.
- Deacon, D., Hewitt, S., Yang, C. M., & Nagata, M. (2000). Event-related potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process. *Cognitive Brain Research*, 9(2), 137-146.
- Deak, G. O. (2000). Hunting the fox of word learning: Why "constraints" fail to capture it. *Developmental Review*, 20(1), 29-80.
- Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science*, 284(5416), 970-974.
- Dollaghan, C. (1985). Child Meets Word: "Fast Mapping" in Preschool Children. *J Speech Hear Res*, 28(3), 449-454.
- Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychol Sci*, 18(1), 35-39.
- Durso, F. T., & Shore, W. J. (1991). Partial knowledge of word meanings. *Journal of Experimental Psychology-General*, 120(2), 190-202.
- Duzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H.-J., & Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences*, 94(11), 5973-5978.
- Eisele, J. A., & Aram, D. M. (1993). Differential effects of early hemisphere damage on lexical comprehension and production. *Aphasiology*, 7(5), 513-523

- Elman, J. (1998). Generalization, simple recurrent networks, and the emergence of structure. *20th annual conference of the Cog Sci society*, 1-13.
- Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science: A Multidisciplinary Journal*, 14(2), 179 - 211.
- Elshout-Mohr, M., & van Daalen-Kapteijns, M. M. (1987). Cognitive processes in learning word meanings. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Evans, K. M., & Federmeier, K. D. (2007). The memory that's right and the memory that's left: Event-related potentials reveal hemispheric asymmetries in the encoding and retention of verbal information. *Neuropsychologia*, 45(8), 1777-1790.
- Federmeier, K. D., & Benjamin, A. S. (2005). Hemispheric asymmetries in the time course of recognition memory. *Psychonomic Bulletin & Review*, 12, 993-998.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4), 469-495.
- Federmeier, K. D., Mai, H., & Kutas, M. (2005). Both Sides Get the Point: Hemispheric Sensitivities to Sentential Constraint. *Memory & Cognition*, 33, 871-886.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., & Thal, D. (1994). Variability in early communicative development. . *Monographs of the Society for Research in Child Development*, 59(Serial no, 242).
- Fodor, J. A. (1983). *The modularity of mind : an essay on faculty psychology*. Cambridge, Mass.: MIT Press.
- Francis, W. S. (1999). Cognitive integration of language and memory in bilinguals: Semantic representation. *Psychological Bulletin*, 125(2), 193-222.
- Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22(5), 1219-1248.
- Friedrich, M., & Friederici, A. D. (2004). N400-like Semantic Incongruity Effect in 19-Month-Olds: Processing Known Words in Picture Contexts. *J. Cogn. Neurosci.*, 16(8), 1465-1477.
- Friedrich, M., & Friederici, A. D. (2005a). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *Neuroreport*, 16(6), 653-656.

- Friedrich, M., & Friederici, A. D. (2005b). Semantic sentence processing reflected in the event-related potentials of one- and two-year-old children. *Neuroreport*, 16(16), 1801-1804.
- Friedrich, M., & Friederici, A. D. (2006). Early N400 development and later language acquisition. *Psychophysiology*, 43(1), 1-12.
- Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, 89(2), 105-132.
- Geschwind, N. (1965). Disconnexion syndromes in animals and man. I. *Brain*, 88(2), 237-294.
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73(2), 135-176.
- Gold, B. T., & Buckner, R. L. (2002). Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. *Neuron*, 35(4), 803-812.
- Goldberg, A. E., Casenhiser, D., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, 14, 289-316.
- Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: rate, content, and the vocabulary spurt. *J Child Lang*, 17(1), 171-183.
- Goldfield, B. A., & Reznick, J. S. (1996). Measuring the vocabulary spurt: a reply to Mervis & Bertrand. *J Child Lang*, 23(1), 241-246.
- Golinkoff, R. M., Hirsh-Pasek, J., Bloom, L., Smith, L. B., Woodward, A. L., Akhtar, N., et al. (2000). *Becoming a word learner: A debate on lexical acquisition*. New York, NY: Oxford University Press.
- Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework. *J Child Lang*, 21(1), 125-155.
- Gopnik, A., Choi, S., & Baumberger, T. (1996). Cross-linguistic differences in early semantic and cognitive development. *Cognitive Development*, 11(2), 197-225.
- Gopnik, A., & Meltzoff, A. N. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development*, 58(6), 1523-1531.
- Gopnik, A., & Meltzoff, A. N. (1992). Categorization and Naming: Basic-Level Sorting in Eighteen-Month-Olds and Its Relation to Language. *Child Development*, 63(5), 1091-1103.

- Gopnik, A., & Meltzoff, A. N. (1993). Words and thoughts in infancy: The specificity hypothesis and categorization and naming. . In C. Rovee-Collier & L. Lipsitt (Eds.), *Advances in infancy reserach*. New Jersey: Ablex.
- Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, theories*. Cambridge, MA: MIT Press.
- Grose-Fifer, J., & Deacon, D. (2004). Priming by natural category membership in the left and right cerebral hemispheres. *Neuropsychologia*, 42(14), 1948-1960.
- Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. M. Brown & P. Hagoort (Eds.), *The Neurocognition of Language* (Vol. 273-316). Oxford: Oxford University Press.
- Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: a test of the critical-period hypothesis for second-language acquisition. *Psychol Sci*, 14(1), 31-38.
- Halsband, U. (2006). Bilingual and multilingual language processing. *Journal of Physiology-Paris*, 99(4-6), 355-369.
- Hart, B., & Risley, T. (1995). *Meaningful Differences in Everyday Parenting and Intellectual Development in Young American Children*. Baltimore, MD: Brookes.
- Heibeck, T. H., & Markman, E. M. (1987). Word Learning in Children: An Examination of Fast Mapping. *Child Development*, 58(4), 1021-1034.
- Hernandez, A. E., Dapretto, M., Mazziotta, J., & Bookheimer, S. (2001). Language switching and language representation in Spanish-English bilinguals: An fMRI study. *Neuroimage*, 14(2), 510-520.
- Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, 95(3), 781-787.
- Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Dev*, 73(2), 418-433.
- Hollich, G. J., Jusczyk, P., & Brent, M. How infants use the words they know to learn new words. 1-12.
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, 10(4), 785-813.
- Huttenlocher, J., Haight, W., Bryk, A., & Seltzer, M. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27(2), 236-248.
- Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. C. (2002). Language input at home and at school: Relation to syntax. *Cognitive Psychology*, 45, 337-374.

- Illes, J., Francis, W. S., Desmond, J. E., Gabrieli, J. D. E., Glover, G. H., Poldrack, R., et al. (1999). Convergent cortical representation of semantic processing in bilinguals. *Brain and Language*, 70(3), 347-363.
- Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning Vocabulary Through Reading. *American Educational Research Journal*, 21(4), 767-787.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(1), 60-99.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory & Language*, 49, 133-156.
- Keibel, H., & Elman, J. (2004). *Distributional regularities in German child-direct speech*. Paper presented at the Architectures and Mechanisms of Language Processing, Aix en Provence, France.
- Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associated thesaurus of English and its computer analysis. In A. J. Aitken, R. Bailey & N. Hamilton-Smith (Eds.), *The computer and literary studies*. Edinburgh: Edinburgh University Press.
- Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5(2), 196-214.
- Koivisto, M. (1997). Time course of semantic activation in the cerebral hemispheres. *Neuropsychologia*, 35(4), 497-504.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12), 463-470.
- Kutas, M., & Federmeier, K. D. (2001). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, 4(12), 463-470.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203-205.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947), 161-163.
- Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory States and Processes* (pp. 217-237). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. (pp. 83-143). San Diego: Academic Press, Inc.
- Lee, H.-S., Fujii, T., Okuda, J., Tsukiura, T., Umetsu, A., Suzuki, M., et al. (2003). Changes in brain activation patterns associated with learning of Korean words by Japanese: an fMRI study. *NeuroImage*, 20(1), 1-11.
- Lenneberg, E. H. (1967). Language in the context of growth and maturation. In *Biological Foundations of Language* (pp. 125-187). New York: John Wiley & Sons.
- Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Netw*, 17(8-9), 1345-1362.
- Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, 7(4), 618-630.
- Luria, A. R. (1973). Neuropsychological studies in the USSR. A review (part II). *Proc Natl Acad Sci U S A*, 70(4), 1278-1283.
- Mandler, J. M. (1996). Preverbal representation and language. In P. Bloom, M. A. Peterson, L. Nadel & M. F. Garrett (Eds.), *Language and Space*. Cambridge, MA: MIT Press.
- Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, 21(2), 339-366.
- Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, 86(1), 70-82.
- Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.
- Markman, E. M. (1992). Constraints on Word Learning - Speculations About Their Nature, Origins, and Domain Specificity. *Minnesota Symposia on Child Psychology*, 25, 59-101.
- Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognit Psychol*, 47(3), 241-275.
- Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, 385(6619), 813-815.
- Marsolek, C. J. (1995). Abstract visual-form representations in the left cerebral hemisphere. *Journal of Experimental Psychology: Human Perception & Performance*, 21(2), 375-386.

- Marsolek, C. J., Kosslyn, S. M., & Squire, L. R. (1992). Form-specific visual priming in the right cerebral hemisphere. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18(3), 492-508.
- Marsolek, C. J., Nicholas, C. D., & Andresen, D. R. (2002). Interhemispheric communication of abstract and specific visual-form information. *Neuropsychologia*, 40(12), 1983-1999.
- Marsolek, C. J., Schacter, D. L., & Nicholas, C. D. (1996). Form-specific visual priming for new associations in the right cerebral hemisphere. *Memory & Cognition*, 24(5), 539-556.
- Marsolek, C. J., Squire, L. R., Kosslyn, S. M., & Lulenski, M. E. (1994). Form-specific explicit and implicit memory in the right cerebral hemisphere. *Neuropsychology*, 8(4), 588-597.
- Mayberry, R. I., & Lock, E. (2003). Age constraints on first versus second language acquisition: evidence for linguistic plasticity and epigenesis. *Brain Lang*, 87(3), 369-384.
- Mayberry, R. I., Lock, E., & Kazmi, H. (2002). Linguistic ability and early language exposure. *Nature*, 417(6884), 38.
- McDonough, L., Choi, S., & Mandler, J. M. (2003). Understanding spatial relations: flexible infants, lexical adults. *Cognit Psychol*, 46(3), 229-259.
- McKeown, M. G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly*, 20, 482-496.
- McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat Neurosci*, 7(7), 703-704.
- McRae, K., deSa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology-General*, 126(2), 99-130.
- McShane, B. (1979). Topic Review: The development of naming. *Linguistics*, 17, 879-905.
- Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., et al. (2004). Neurolinguistics: Structural plasticity in the bilingual brain. *Nature*, 431(7010), 757-757.
- Mervis, C. B. (1983). Acquisition of a lexicon. *Contemporary Educational Psychology*, 8, 210-236.
- Mervis, C. B., & Bertrand, J. (1994). Acquisition of the Novel Name^Nameless Category (N3C) principle. *Child Development*, 65(6), 1646-1662.
- Mervis, C. B., & Bertrand, J. (1995). Early lexical acquisition and the vocabulary spurt: A response to Goldfield & Reznick. *Journal of Child Language*, 22(2), 461-468.

- Mestres-Misse, A., Rodriguez-Fornells, A., & Munte, T. F. (2006). Watching the Brain during Meaning Acquisition. *Cereb. Cortex*, bhl094.
- Meyer, D. E. (1970). On the representation and retrieval of stored semantic information. *Cognitive Psychology*, 1(3), 242-299.
- Mills, D. L., Coffey-Corina, S., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology*, 13(3), 397-445.
- Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*, 5(3), 317-334.
- Mills, D. L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: effects of vocabulary size and experience. *Cognitive Development*, 20(1), 19-31.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychol Rev*, 92(3), 289-316.
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning Word Meanings from Context during Normal Reading. *American Educational Research Journal*, 24(2), 237-270.
- Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning Words from Context. *Reading Research Quarterly*, 20(2), 233-253.
- Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Developmental Science*, 6, 136-142.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In E. Derek Besner & E. Glyn W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition*. (pp. 264-336): Hillsdale, NJ, US Hillsdale, NJ, US.
- Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, 38(1-2, Serial No 149), 136.
- Newport, E. L. (1997). Motherese: The speech of mothers to young children. . In N. J. Castellan, D. B. Pisoni & G. R. Potts (Eds.), *Cognitive Theory* (Vol. 2, pp. 177-217). Hillsdale, New Jersey: Erlbaum.
- Nobre, A. C., & McCarthy, G. (1994). Language-Related Erps - Scalp Distributions and Modulation by Word Type and Semantic Priming. *Journal of Cognitive Neuroscience*, 6(3), 233-255.
- Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP Study of Second Language Learning after Childhood: Effects of Proficiency. *J. Cogn. Neurosci.*, 17(8), 1212-1228.

- Olguin, R., & Tomasello, M. (1993). Twenty-Five-Month-Old Children Do Not Have a Grammatical Category of Verb. *Cognitive Development*, 8(3), 245-272.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory & Language*, 31(6), 785-806.
- Osterhout, L., McLaughlin, J., Pitkanen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice Learners, Longitudinal Designs, and Event-Related Potentials: A Means for Exploring the Neurocognition of Second Language Processing. *Language Learning*, 56(s1), 199-230.
- Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Dev*, 76(4), 763-782.
- Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *J Exp Psychol Learn Mem Cogn*, 31(6), 1281-1292.
- Piaget, J. (1971). *Langage et la pensÃ©e chez l'enfant*. London : New York :: Routledge & Kegan Paul; Humanities Press.
- Pinker, S. (1991). Rules of language. *Science*, 253(5019).
- Pinker, S. (1994). *The language instinct: The new science of language and mind*. . London: Penguin.
- Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review*, 107(4), 786-823.
- Plunkett, K., & Elman, J. L. (1997). *A handbook for connectionist simulations*. Cambridge, MA: MIT Press.
- Plunkett, K., Sinha, C., Moller, M. F., & Strandsby, O. (1992). Symbol Grounding or the Emergence of Symbols? Vocabulary Growth in Children and a Connectionist Net. *Connection Science*, 4(3), 293 - 312.
- Plunkett, K., Sinha, C., Moller, M. F., & Stransby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, 4, 293-312.
- Poulin-Dubois, D., Graham, S., & Sippola, L. (1995). Early lexical development: the contribution of parental labelling and infants' categorization abilities. *J Child Lang*, 22(2), 325-343.

- Raboyeau, G., Marie, N., Balduyck, S., Gros, H., Démonet, J.-F., & Cardebat, D. (2004). Lexical learning of the English language: a PET study in healthy French subjects. *NeuroImage*, 22(4), 1808-1818.
- Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology*, 55(A), 1339-1362.
- Rehder, B., & Ross, B. H. (2001). Abstract coherent categories. *J Exp Psychol Learn Mem Cogn*, 27(5), 1261-1275.
- Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental changes in mental arithmetic: evidence for increased functional specialization in the left inferior parietal cortex. *Cereb Cortex*, 15(11), 1779-1790.
- Rogers, T. T., & McClelland, J. L. (2006). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
- Rohde, D. L. T. (1999). *The simple language generator: Encoding complex languages with simple grammars*. (Technical Report). Pittsburgh, PA: Carnegie Mellon, Department of Computer Science.
- Ruz, M., Madrid, E., Lupianez, J., & Tudela, P. (2003). High density ERP indices of conscious and unconscious semantic priming. *Cognitive Brain Research*, 17(3), 719-731.
- Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, 24(2), 232-252.
- Schwanenflugel, P. J., Stahl, S. A., & McFalls, E. L. (1997). Partial word knowledge and vocabulary growth during reading comprehension. *Journal of Literacy Research*, 29(4), 531-553.
- Singleton, D. M. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- Smith, L., Jones, S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? , 1-55.
- Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object Name Learning Provides On-the-Job Training for Attention. *Psychological Science*, 13(1), 13-19.
- Snow, C. E., Burns, M. S., & Griffin, P. (1998). Preventing Reading Difficulties in Young Children.
- Snow, C. E., & Ferguson, C. A. (1977). *Talking to children: Language input and acquisition*. Cambridge: Cambridge University Press.

- Stahl, S. A., & Fairbanks, M. M. (2006). The effects of vocabulary instruction: A model-based meta-analysis. In D.-S. K. A. & M. C. McKenna (Eds.), *Reading research at work: Foundations of effective practice*. New York: The Guilford Press.
- Stein, M., Dierks, T., Brandeis, D., Wirth, M., Strik, W., & Koenig, T. (2006). Plasticity in the adult language system: A longitudinal electrophysiological study on second language learning. *NeuroImage*, 33(2), 774-783.
- Stemmer, B., Giroux, F., & Joannette, Y. (1994). Production and Evaluation of Requests by Right Hemisphere Brain-Damaged Individuals. *Brain and Language*, 47(1), 1-31.
- Sternberg, R. J. (1987). Most vocabulary is learned from context. . In M. G. McKeown & M. E. Curtis (Eds.), *The Nature of Vocabulary Acquisition* (pp. 89-106). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Sugarman, S. (1981). The cognitive basis of classification in very young children: an analysis of object-ordering trends. *Child Dev*, 52(4), 1172-1178.
- Tan, L. H., Spinks, J. A., Feng, C. M., Siok, W. T., Perfetti, C. A., Xiong, J. H., et al. (2003). Neural systems of second language reading are shaped by native language. *Human Brain Mapping*, 18(3), 158-166.
- Thal, D. J., Marchman, V. A., Stiles, J., Aram, D., Trauner, D., Nass, R., et al. (1991). Early lexical development in children with focal brain injury. *Brain & Language*, 40(4), 491-527.
- Thomas, H., & Dahlin, M. P. (2000). Inferring children's categorizations from sequential touching behaviors: an analytical model. *Psychol Rev*, 107(1), 182-194.
- Thompson-Schill, S. L., Swick, D., Farah, M. J., D'Esposito, M., Kan, I. P., & Knight, R. T. (1998). Verb generation in patients with focal frontal lesions: a neuropsychological test of neuroimaging findings. *Proc Natl Acad Sci U S A*, 95(26), 15855-15860.
- Tomasello, M. (1996). Young children's productivity with word order and verb morphology. 1-40.
- Tomasello, M., & Olguin, R. (1993). Twenty-Three-Month-Old Children Have a Grammatical Category of Noun. *Cognitive Development*, 8(4), 451-464.
- Tompkins, C. A., Scharp, V. L., Meigh, K. M., & Fassbinder, W. (2008). Coarse coding and discourse comprehension in adults with right hemisphere brain damage. *Aphasiology*, 22(2), 204 - 223.
- Tomasello, M. (2000). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

- Torkildsen, J. v. K., Sannerud, T., Syversen, G., Thormodsen, R., Simonsen, H. G., Moen, I., et al. (2006). Semantic organization of basic-level words in 20-month-olds: An ERP study. *Journal of Neurolinguistics*, 19(6), 431-454.
- Torkildsen, J. v. K., Syversen, G., Simonsen, H. G., Moen, I., & Lindgren, M. (2007). Electrophysiological correlates of auditory semantic priming in 24-month-olds. *Journal of Neurolinguistics*, 20(4), 332-351.
- Van Lancker, D. R. K., Daniel. (1987). Comprehension of familiar phrases by left but not right hemisphere damaged patients. *Brain & Language*, 32(2), 265-277.
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, 18(4).
- Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, 3(2), 131-150.
- Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: context, relational properties, and concept naturalness. *Cognit Psychol*, 18(2), 158-194.
- Waxman, S. R., & Booth, A. E. (2000). Principles that are invoked in the acquisition of words, but not facts. *Cognition*, 77(2), B33-B43.
- Weems, S. A., & Zaidel, E. (2005). Repetition priming within and between the two cerebral hemispheres. *Brain and Language*, 93(3), 298-307.
- Weizman, Z. O., & Snow, C. E. (2001). Lexical Input as Related to Children's Vocabulary Acquisition: Effects of Sophisticated Exposure and Support for Meaning. *Developmental Psychology*, 37(2), 265-279.
- Wells, C. G. (1981). *Learning through interaction: The study of language development*. (Vol. Cambridge University Press.): Cambridge.
- Weylman, S. T., Brownell, H., Roman, M., & Gardner, H. (1989). Appreciation of indirect requests by left- and right-brain-damaged patients: The effects of verbal context and conventionality of wording. *Brain & Language*, 36(4), 580-591.
- Whorf, B. L. (1956). *Language, Thought and Reality*. Cambridge, MA: MIT Press.
- Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*, 20(1), 6-11
- Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *J Exp Psychol Learn Mem Cogn*, 21(2), 449-468.

- Wittgenstein, L. (1953). *Philosophical Investigations*. New York: Macmillan.
- Woodward, A. L., & Markman, E. M. (1998). Early word learning. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Vol. II. Cognition, perception, and language*. . New York: John Wiley and Sons.
- Zavrel, J. (1996a). Lexical space: Learning and using continuous linguistic representations: Utrecht University, Utrecht, Netherlands.
- Zavrel, J. (1996b). Lexical space: Learning and using continuous linguistic representations.: Utrecht University, Utrecht, Netherlands.