The Longitudinal Development of Lexical Network Knowledge in L2 Learners: Multiple Methods and Parallel Data

THE LONGITUDINAL DEVELOPMENT OF LEXICAL NETWORK KNOWLEDGE IN L2 LEARNERS: MULTIPLE METHODS AND PARALLEL DATA

by

CYNTHIA BERGER

Under the Direction of Scott Crossley, PhD

ABSTRACT

This dissertation analyzes the development of adult second language (L2) English learners' depth of lexical knowledge over six months through a series of three longitudinal experimental studies. The goal of the project is to provide a better understanding of how the English lexicon develops over time in L2 learners. Methods employed include vocabulary size assessment, a word association task, lexical decision semantic priming, and the computational analysis of subjects' spoken lexical output. Results found little evidence of longitudinal development in L2 subjects' lexical network knowledge, with the exception of L2 word associations, which actually became less native-like over time. In addition, results indicated no observed relationship between vocabulary size and the dependent measures obtained in the three studies. This dissertation project has theoretical implications for our understanding of depth of lexical knowledge and its rate of development. The project also has methodological implications for future experimental and/or longitudinal investigations of the L2 lexicon.

INDEX WORDS: Vocabulary, Second language, English as a Second Language, Language learning, Mental lexicon, Depth of lexical knowledge

THE LONGITUDINAL DEVELOPMENT OF LEXICAL NETWORK KNOWLEDGE IN L2 LEARNERS: MULTIPLE METHODS AND PARALLEL DATA

by

CYNTHIA BERGER

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2018

THE LONGITUDINAL DEVELOPMENT OF LEXICAL NETWORK KNOWLEDGE IN L2 LEARNERS: MULTIPLE METHODS AND PARALLEL DATA

by

CYNTHIA BERGER

Committee Chair: Scott Crossley

Committee: Scott Jarvis

YouJin Kim

Ute Römer

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2018

DEDICATION

To my family. You are love and intensity. Without the former, I would never have survived. Without having inherited the latter, I might never have finished.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of a number of people. First, I am deeply indebted to my advisor and friend Scott Crossley. Scott has spent countless hours guiding me through statistical analyses, existential crises, career milestones, and the iterative process of completing this project. If it weren't for Scott, I would not be the researcher I am today, and I'm not certain I would have made it through the last five years. Thank you.

YouJin Kim taught me everything I know about second language acquisition, and her positive encouragement over the years has impacted me in ways she will likely never realize. In addition to travel adventures, chocolate, and friendship, I wish to thank Ute Römer for her unparalleled expertise in corpus linguistics and usage-based language learning. I would also like to thank Scott Jarvis, who took an interest in my early research and whose feedback improved this project in a number of ways.

On the whole, I am grateful to the faculty and staff of the Department of Applied Linguistics and ESL, as well as to the Intensive English Program instructors (IEP) who demonstrate daily the value of language practitioner knowledge. I am equally indebted to the IEP students and undergraduates who participated in these studies, and to Duolingo for giving me the flexibility and encouragement to finish this dissertation. Finally, I wish to thank my dear friends and graduate school colleagues.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EFL	English as a Foreign Language
ESL	English as a Second Language
FB	Functionally bilingual
IEP	Intensive English Program
L1	First language
L2	Second language
LD	Lexical decision
LES	Language Experience Survey
LME	Linear mixed effects (model)
NES	Native English speaker
NLP	Natural language processing
NNS	Non-native speakers
RTs	Response times
SLA	Second language acquisition
SPE	Semantic priming effects
UG	Undergraduate
VLT	Vocabulary Levels Test
WA	Word association

1 INTRODUCTION

Empirical evidence supports a strong relationship between vocabulary and performance across a wide range of second language (L2) proficiency (Alderson, 2005; Daller, Van Hout, & Treffers-Daller, 2003; David, 2008; Laufer, 1992; Meara, 1996; Read, 2000, 2004; Schmitt, 2010), indicating that vocabulary knowledge is an essential component of competence in an L2. Nevertheless, lexical knowledge remains an “inherently complex” construct (Read, 2000, p. 93), because the lexical acquisition system is still considered “too complex and variable” for simple description (Schmitt, 2010, p. 38). For these reasons, second language acquisition (SLA) researchers have yet to identify the key mechanisms and patterns of L2 lexical development. This poses problems for a wide range of researchers who wish to better assess lexical knowledge and its development over time but must first determine how best to operationalize the construct of lexical knowledge itself and second, how to measure it.

1.1 Lexical Knowledge

The term *lexical knowledge* refers to an underlying mental representation encoded in long-term memory that contains the known words of a language (Barcroft, Schmitt, & Sunderman, 2011, p. 571; Bialystok, 1994; Dóczi & Kormos, 2016). Most linguists agree that the fundamental aspect of such knowledge is the association of a word’s form (i.e., its sound or spelling) with its basic meaning (Barcroft et al., 2011; Nation, 1990; Read, 2000). Traditionally, vocabulary researchers have attempted to highlight the distinction between early form-meaning knowledge and more nuanced lexical knowledge by referring to *breadth* vs. *depth* of lexical knowledge (Anderson & Freebody, 1981).

Breadth of vocabulary size has traditionally been associated with the goal of arriving at an estimation of the vocabulary size of a language speaker (e.g., Hazenberg & Hulstijn, 1996;

Meara & Buxton, 1987; Nurweni & Read, 1999). A valid estimate of a speaker's vocabulary size is useful because it allows researchers to better understand how lexical knowledge develops over time or how many form-meaning pairings a speaker needs to know in order to complete a particular task or participate in certain interactions (e.g., Nation, 2006). Meanwhile, depth of knowledge is often attributed to the quality of vocabulary knowledge, rather than to its size (Nation, 2001; Read, 1993, 2004). While we know a great deal about how the form-meaning component of word knowledge is acquired, we know significantly less about the development of depth of vocabulary knowledge, particularly over time (Dóczi & Kormos, 2016; Schmitt, 2010). So-called *word-centered* investigations of depth of knowledge typically take a developmental or a dimensions-based approach (Schmitt, 2010). While a developmental approach (e.g., Read, 2000) is concerned with the incremental acquisition of lexical items (i.e., words), a dimensions-based approach seeks to determine the various components of word knowledge one might have about a particular lexical item (e.g., its grammatical function, spelling, prefixes or suffixes that may form the word; cf. Nation, 2001).

More recently, researchers have begun to acknowledge the role of lexical organization in conceptions of depth of vocabulary knowledge (e.g., Read, 2004). As opposed to the word-centered approach, this focus is referred to as a *lexicon-based* perspective (Schmitt, 2010) and is concerned with speakers' knowledge of lexical items' relationships to one another (Aitchison, 1994; Singleton, 1999, 2000). Under a lexicon-based approach, lexical development is conceived as the growth and restructuring of an entire *system*, rather than the accrual of individual lexical *items* (Henriksen, 1999). The assumption that lexical knowledge is cognitively organized in a structured way is relatively unchallenged, and a growing body of scholars insist that contemporary approaches to understanding human language include the development of network knowledge and the ability to trace and measure the emergence of connections between words "on a more symbolic level" (Read, 2000, p. 123).

Unfortunately, researchers have tended to neglect speakers' lexical networks, focusing instead on the method by which initial word meanings are mapped onto form (i.e., breadth) or on the quality with which individual words are known (i.e., the dimensions-based approach to depth) (cf Read, 2000). One reason for this is because the process of network building is slower and more gradual than the relatively fast process of mapping meaning onto form (Greidanus & Nienhuis, 2001; Henriksen, 1999). Longitudinal studies, with multiple data collection points across a range of task types, would be needed in order to adequately observe network building and derive descriptive models (Henriksen, 1999, 2008). While there is already a scarcity of longitudinal studies in SLA (Ortega & Iberri-Shea, 2005), this is especially true for research that investigates the development of the mental lexicon (Dóczi & Kormos, 2016; Henriksen, 1999). To date, no longitudinal study exists which would allow researchers to compare multiple methods for measuring L2 lexical network knowledge over time across the same group of subjects.

Another reason there has been little research investigating L2 lexical network development is because the mental lexicon remains "inherently ill-defined, multidimensional, variable and thus resistant to neat classification" (Dóczi & Kormos, 2016, p. 224): While there are a range of existing tools to measure vocabulary size and range, there is little consensus on how best to go about measuring and assessing the network knowledge of language speakers (i.e., the types of links in the mental lexicon and its structural properties).

The current project addresses these gaps in knowledge and available data by offering a longitudinal investigation of L2 depth of lexical knowledge from a network-based perspective.

This dissertation reports on a 6-month study which observed the organization and longitudinal development of English lexical network knowledge in adult L2 English learners using parallel data derived from three experimental methods: word associations (Study 1), lexical decision semantic priming (Study 2), and spoken lexical production (Study 3). For these three related studies, data were collected from both L2 English learners and highly proficient English monolinguals and bilinguals. Longitudinal data were only collected from L2 learners. Findings from the three inter-related studies may provide important contributions to theoretical understandings of how L2 English lexicons are organized, to empirical evidence of how L2 lexicons develop over time, and to the methodological practices of SLA researchers who investigate lexical knowledge.

This dissertation is guided by the following questions:

1. What patterns of development are revealed by the longitudinal analysis of L2 learners' English word associations with relation to time and vocabulary size?
2. What patterns of development are revealed by the longitudinal analysis of L2 learners' English semantic priming with relation to time and vocabulary size?
3. What patterns of development are revealed by the longitudinal analysis of associational lexical knowledge in L2 learners' spoken production with relation to time and vocabulary size?

This dissertation manuscript is organized as follows. First, I review the relevant literature concerning lexical knowledge, including both how the construct is defined and previous attempts at measuring lexical knowledge (Chapter 2). Next, I provide a brief overview of the three focal studies that comprise this dissertation project and the research subjects who participated in data collection (Chapter 3). The next three chapters (Chapters 4-6) are each dedicated to a separate focal study that investigates the development of L2 lexical development according to the research questions above, and the final chapter (Chapter 7) summarizes results and discusses implications.

2 LEXICAL KNOWLEDGE

This chapter takes as its starting point the traditional distinction between breadth and depth of L2 lexical knowledge. Using this paradigm as a framework, I review the manner in which lexical knowledge has been operationalized and assessed by L2 researchers representing a variety of approaches to conceptualizing lexical knowledge. Because of the unique ability for longitudinal studies to offer insights into SLA as a whole and to L2 lexical development specifically, I highlight work that has measured the breadth and depth of L2 lexical knowledge and their development over time. Next, I acknowledge other constructs that have been used to conceptualize L2 lexical knowledge and discuss them in relation to the breadth vs. depth paradigm. I conclude this chapter by outlining existing research gaps and future directions for researchers investigating the development of L2 lexical knowledge, with particular emphasis on the potential for psycholinguistic investigations of lexical knowledge to capture more nuanced development of L2 knowledge overtime. Crucially, throughout this chapter, focus is given to the development of L2 lexical knowledge as a *product* (Leow, 2015, p. 50), rather than examining the conditions of learning that might result in different types of knowledge (e.g., incidental learning, Laufer & Hulstijn, 2001).

2.1 Defining Constructs

2.1.1 Words

Historically, the manner in which words have been defined typically align with one of three perspective: formal, semantic, or psycholinguistic (Dóczi & Kormos, 2016).

2.1.1.1 Formal

In the formal approach, a string of letters separated by spaces orthographically is considered a word. This is the most intuitive definition and one that is likely to make the most sense to non-specialists, as well as to many language teachers and students, for whom conceiving of words in this manner is both familiar and convenient (Schmitt, 2010). It is an equally convenient operationalization for L2 researchers interested in counting or extracting words automatically using word processors or natural language processing tools. However, such a definition is problematic to the extent that it ignores words' semantic meaning and thus risks subsuming polysemous and homonymous words into one unit (Barcroft, Schmitt, & Sunderman, 2011; Dóczi & Kormos, 2016). For similar reasons, this approach makes it difficult to quantify the actual words learners know (Bogaards, 2001; Dóczi & Kormos, 2016).

2.1.1.2 Semantic

The semantic approach to word definition is primarily concerned with word meaning. Operating from this perspective, some researchers (Laufer & Nation, 1995) have argued that a basic word form and all of its derivations and inflections (e.g., *write, writes, writing, writer*) constitute one unit, the word family. Still, the semantic approach is complicated by the ubiquity of function words, such as prepositions, which make frequent appearances in language use despite having little to no semantic or pragmatic meaning. Furthermore, function words are often used in a string of words that takes on a meaning quite different from the meaning of its constituent words (Dóczi & Kormos, 2016). For this reason, others have taken to operationalizing vocabulary as consisting not of words per se, but of *lexical units* (e.g., Cruse, 1986). According to this lexicosemantic approach, a *lexical unit* is defined as the union of at least one traditionally recognized orthographic word and a single semantic meaning. A great deal of work has been conducted in the last 30 years exploring the role that multiple-word lexical units play in language as a whole and in lexical knowledge specifically (Biber & Conrad, 1999;

Granger & Meunier, 2008; Moon, 1997; Nattinger & DeCarrico, 1992; Schmitt & Carter, 2004; Wray, 2002).

2.1.1.3 Psycholinguistic

The psycholinguistic approach "considers how users of a language or multiple languages store and retrieve words from their mental lexicon" (Dóczi & Kormos, 2016, p. 5). From the psycholinguistic perspective, the basic unit of storage and representation is the *lemma*. A lemma is typically defined as the base form of a word and all of its inflected forms within the same part of speech (Kucera, 1982). Such a definition usually treats derivative forms as separate lemmas (e.g., *writings* and *writer* would be considered two separate lemmas*)*.

Like certain versions of the lexicosemantic approach, the psycholinguistic approach to word definition allows for recurring sequences of words to be stored and retrieved as a single unit (Wray, 2002). For example, Schmitt (2010) uses the term *formulaic language* to refer to the phenomenon in language whereby multi-word lexical items may be stored and processed holistically in the mind. He distinguishes between formulaic language and vocabulary by using the term *formulaic sequence* to refer to a multi-word lexical item that is stored holistically, while *lexical item* includes both individual words and formulaic sequences.

2.1.2 Lexical Knowledge

While Harley (2008) suggests that the definition of a word proves to be "a somewhat slippery customer" (p. 6), how to go about characterizing lexical knowledge as a whole turns out to be an even slipperier enterprise. The term *lexis* is typically used to refer "to all the words in a language, the entire vocabulary of a language" (Barcroft et al., 2011, p. 571). As for *knowledge* itself, Dóczi and Kormos (2016) reference cognitive approaches (e.g., Bialystok, 1994) when defining *knowledge* as "an underlying mental representation encoded in long-term memory" (p.

6). Thus, *lexical knowledge* refers to an underlying mental representation, encoded in long-term memory, that contains all the words of a language. Most researchers agree that the most fundamental aspect of such knowledge is the association of a word's form (phonological or orthographical) with its basic meaning (Barcroft et al., 2011; Nation, 1990; Read, 2000), though even this basic level of knowledge has been demonstrated to be incremental and involve nuances of understanding (Henriksen, 1999; Read, 2004).

The term *vocabulary*, though intuitively useful, is not without controversy. While *vocabulary* and *lexis* are used interchangeably by many L2 researchers and practitioners, others (Paradis, 2009) assert that the two are subserved by entirely different memory systems. Even among those who assume that *vocabulary* and *lexis* refer to roughly the same underlying mental representation, *vocabulary* tends to be preferred by researchers interested in language pedagogy and assessment, while *lexis* is used by researchers who investigate language learning and representation in a slightly more theoretical manner. Finally, cognitive linguists and psycholinguists often use the term *lexicon* to refer to the *mental lexicon*, the "mental dictionary" (Harley, 2008, p. 7) in which all of our information about the *lexis* is organized and stored (the mental lexicon will be discussed in greater detail below).[1]

Importantly, there is no comprehensive theory of L2 vocabulary acquisition. Schmitt (2010) explains that while the field has made progress in understanding how some isolated aspects of vocabulary develop, "the overall acquisition system is far too complex and variable for us to comprehend in its entirety, and so it still eludes description" (p. 38). Of course, any attempt to explain or predict the acquisition of L2 vocabulary knowledge (or even to define the

---

[1] In this chapter, *lexical knowledge* and *vocabulary knowledge* will be used interchangeably, unless the paradigm under discussion differentiates the terms to refer to distinct knowledge systems (e.g., Paradis, 2009), in which case this distinction will be explicitly acknowledged.

construct itself) will be informed by assumptions about the relationship between lexical and syntactic encoding. There are two distinct predominant theoretical perspectives as far as this issue is concerned. In the first, it is argued that lexis and grammar are completely separable (Pinker, 1991). This is the position largely taken by generativist and formalist linguists (e.g., Chomsky, 1965, 1981; Keenan & Comrie, 1977; Pollock, 1989). From the second perspective, lexis and grammar are assumed to be inter-related, if not a unitary continuum (Gries & Stefanowitsch, 2007; Römer, 2009; Stemberger & MacWhinney, 1988). Among the approaches to L2 learning that take this second perspective is usage-based learning (Bybee, 2008; N. C. Ellis, 2002b, 2008; Goldberg, 2002, 2006; Tomasello, 2003, 2007; Tyler, 2010), which argues, among other things, that words play a central role in language acquisition and that all linguistic units, from traditional orthographic words to syntactic structures, have a form-meaning mapping. That said, Barcroft et al. (2011) note that even generative and formalist linguists, while continuing to maintain the dichotomy between lexis and syntax, have begun to acknowledge the role of words and lexicalized phrases in constraining syntactic structure.

In the current chapter, these divergent theoretical perspectives are not directly addressed. Instead, this chapter takes as its starting point the manner in which L2 vocabulary knowledge has traditionally been operationalized and assessed within L2 pedagogical research and assessment practice, both of which have historically maintained a distinction (functional, if not theoretical) between grammar and vocabulary. That said, the approach taken here upholds the usage-based claims that language learning is frequency-driven (e.g., N. C. Ellis, 2002a, 2002b) and that meaning in language extends to forms beyond the single world in isolation (Cowie, 1998; Granger & Meunier, 2008; Granger & Paquot, 2008; Howarth, 1998; Römer, 2009).

2.1.3 Breadth vs. depth of knowledge

The distinction between *breadth* and *depth* of vocabulary knowledge has been conceptualized in a variety of ways over the last 35 years. This dichotomizing of knowledge is first attributed to Anderson and Freebody (1981), who clearly distinguished between the two types of knowledge:

> The first may be called "breadth" of knowledge, by which we mean the number of words for which the person knows at least some of the significant aspects of meaning. ... [There] is a second dimension of vocabulary knowledge, namely the quality or "depth" of understanding. We shall assume that, for most purposes, a person has a sufficiently deep understanding of a word if it conveys to him or her all of the distinctions that would be understood by an ordinary adult under normal circumstances. (p. 92)

Breadth is often associated with the size of a learner's vocabulary while *depth* is assumed to capture the quality of knowledge obtained for given words. These two aspects of vocabulary are explained in greater detail in sections 2.2 and 2.3.

It is important to acknowledge that there have been a variety of attempts to characterize the components of L2 vocabulary knowledge, some of which align with or are subsumed by the traditional breadth vs. depth distinction, while others break way from it completely. For example, Nation (1990) suggested several components of word knowledge, each related to a word's form, function, or meaning with respect to receptive or productive mastery. Chapelle (1994) proposed four dimensions of vocabulary knowledge, including vocabulary size, knowledge of word characteristics, lexicon organization, and processing ability. She conceived of vocabulary knowledge as one of several components comprising the more expansive construct of vocabulary ability. Meanwhile, Henriksen (1999) proposed a three-dimensional global framework of lexical competence whereby lexical knowledge was organized according to the following dimensions: *partial and precise knowledge* (similar to breadth or quantity), *depth of knowledge* (similar to quality of knowledge), and *receptive-productive mastery*. To the extent that is possible, throughout this chapter, I include mention of other approaches to conceptualizing lexical knowledge with reference to where they best belong within the traditional breadth vs. depth framework.

2.1.4 The mental lexicon

Dóczi and Kormos (2016) outline two cognitive approaches toward conceptualizing lexical knowledge. In the first, the underlying mental representations are “person-internal” and largely unrelated and isolated from existing representations (e.g., Chapelle, 1998); in other words, vocabulary knowledge is an “abstract individual trait of learners” (Dóczi & Kormos, 2016, p. 6). This is the approach to conceiving of lexical knowledge that is representative of most L2 vocabulary research. In the second approach, knowledge is viewed as an inter-related network of memory traces, with each link between nodes in the network having variable strength. This manner of conceptualizing vocabulary knowledge motivates the *mental lexicon* approach to L2 vocabulary. Investigations of the mental lexicon are primarily concerned with how language is stored, retrieved, and organized structurally in the mind (Randall, 2007). Henriksen (1999) suggests that while vocabulary research has traditionally focused on the method by which learners map word meaning onto form, as a field, we have "tended to disregard the learners’ ongoing process of constructing and reorganizing their interlanguage semantic networks” (p. 307).

Dóczi and Kormos (2016) suggest that the increasingly blurred boundary between lexis and grammar has resulted in more expansive conceptualizations of the mental lexicon, as it is gradually assumed to contain more and more information. For this reason, and others, Read (2004) warn that the mental lexicon is “inherently ill-defined, multidimensional, variable and thus resistant to neat classification” (p. 224). Randall (2007) suggests that both symbolist and connectionist theoretical perspectives probably explain some characteristics of the mental lexicon and that “learning and teaching approaches will need to encourage both the establishment of connections between words and of the grouping of words on a more symbolic level” (p. 123). Furthermore, Dóczi and Kormos note that while the mental lexicon may not exist as a separate linguistic module or area in the brain (cf. Elman, 2009), the assumption that lexical knowledge is organized in a structured way remains unquestioned and still serves “as a useful metaphor to help us understand how words are stored, retrieved, and learned” (p. 12).

2.1.5 Receptive vs. productive knowledge

An important distinction that’s been alluded to in this chapter but not specifically addressed until now is the distinction between receptive or productive lexical knowledge. The vast majority of research suggests that most learners have more receptive knowledge than productive (Fan, 2000; Laufer, 2005; Laufer & Paribakht, 1998; Melka, 1997). Still, few agree on the ratio between receptive to productive knowledge, with some estimates suggesting that as little as 16% of receptive vocabulary is known productively (e.g., Laufer, 2005a) while others suggest that as much as 75% of receptive vocabulary may be known productively (e.g., Fan, 2000).

The discrepancy in findings regarding receptive vs. productive knowledge is likely due to differences in construct definition and measurement type (Schmitt, 2010). For example, the two knowledge types may be conceptualized with reference to whether a learner is required to show word-level or discourse-level lexical mastery. At the word level, *recognition* occurs when an examinee is presented with the target word and asked to demonstrate that they understand its meaning, while *recall* occurs when examinees are presented with a stimulus of some sort and required to produce the target word from memory (Read, 2000). Regardless, both assess the form-meaning link of the lexical item. At the discourse-level, this same distinction is referred to as *comprehension* and *use*. Thus, researchers may differentiate between *recognition* (word-level mastery) and *comprehension* (discourse-level mastery) when discussing receptive knowledge, or between *recall* (word-level) and actual *use* (discourse-level) when discussing productive knowledge (Schmitt, 2010). For obvious reasons, these terms are often either confounded or used indiscriminately by researchers, resulting in issues with comparability across studies and the reporting and interpretation of scores (Schmitt, 2010).

2.2 Breadth of Lexical Knowledge

Breadth of vocabulary size in L2 research has traditionally been associated with estimation of the vocabulary size of a learner (e.g., Hazenberg & Hulstun, 1996; Meara & Buxton, 1987; Nurweni & Read, 1999). The goal has been to arrive at a principled estimate of the number of words a learner knows, information that is useful to both researchers and language pedagogy practitioners alike. For example, a valid estimate of learners' vocabulary size can allow researchers to better understand how lexical knowledge develops over time or how many words an L2 speaker needs to know in order to complete a particular task or participate in certain L2 interactions (e.g., Nation, 2006).

Due to the purpose of such vocabulary size estimates—and the relatively large number of test items required to arrive at such an estimate—breadth of knowledge has primarily been conceptualized with reference to learners' basic knowledge of the relationship between a lexical item's form (phonological or orthographic) and its core meaning. Tasks used to capture the acquisition of form-meaning mappings usually include checklists, single-word translations, or matching terms to their definitions or synonyms (Read, 2004). A number of test instruments have been developed to measure breadth of vocabulary. Many of the most well-known breadth of vocabulary tests are compiled in Table 2.1. Due to space limitations, only a handful of them are discussed below.

One of the best-known and most widely used vocabulary measures (Meara, 1996; Read, 2005; Schmitt, 2010) to date is the *Vocabulary Levels Test* (*VLT*). The *VLT* measures learners' basic knowledge of a word's form-meaning relationship at four different frequency levels (the most frequent 2K, 3K, 5K, and 10K words). The original format of the *VLT* was developed by Nation (1983) but later modified by Laufer and Nation (1999), Beglar and Hunt (1999), and N. Schmitt, D. Schmitt, and C. Clapham (2001a) (see Table 2.1). Both a receptive and a productive version of the *VLT* exist (see *PVLT* below for discussion of the productive version). The receptive format of the *VLT* involves a form-recognition matching task in which test takers are asked to select the appropriate word from six options to match each of three brief definitions (i.e., stems). Schmitt (2010) points out that while the *VLT* does not actually produce an estimate of a learner's overall vocabulary size, it does offer a frequency profile which can be useful for placement and diagnostic purposes. In a validity study of the *VLT,* Schmitt et al. (2001a) concluded that while the *VLT* was indeed a valid measure of breadth of knowledge, it should only be interpreted as "an indication of whether examinees have an initial knowledge of the most frequent meaning sense of each word in the test" (p. 62).

Table 2.1 Instruments to Measure Breadth of Vocabulary

Name	Mastery + Format	Reference	Description	Comments
Peabody Picture Vocabulary Test	Receptive Meaning-recognition	Pearson (2007)	Subjects listen to word read aloud and then point to picture that represents meaning	Ideal for the very young or the elderly; primarily used by L1 researchers; commercially available
Vocabulary Levels Test (VLT)	Receptive Form-recognition	Nation (1983, 1990), Belar & Hunt (1999), Schmitt et al. (2001)	Stem is definition and options are target words; all targets and distractors taken from the same frequency band; four frequency levels represented (2K, 3K, 5K, and 10K)	Offers a profile of learners' vocabulary rather than overall size; useful for placement and diagnostic purposes
Vocabulary Size Test (VST)	Receptive Meaning-recognition	Nation & Gu (2007), Nation (2008)	Four-option multiple choice format w/ target word in an example sentence as the stem; organized into 1,000-word frequency bands	Designed as test of overall vocabulary size (vs. profile); Interactive format available via *Compleat Lexical Tutor* (http://www.lextutor.ca/tests)
Yes/No Tests (checklists)	Receptive Meaning-recall	Meara (1992) and colleagues	Subjects read lists of lexical items and indicate whether they know an item or not; may contain nonwords to adjust for overly confident examinees	Easy to take; yields higher sample rates; Interactive checklist by Meara (1992) available on *Lextutor* website; *Y_Lex* available for download on Meara's lognostics website (www.lognostics.co.uk)
Computer Adaptive Test of Size and Strength (CATSS)	Receptive (somewhat productive as well) Combination of formats	Laufer & Goldstein (2004)	Four item formats based on combination of recognition-recall and form-meaning links	Computerized; unlike static tests (above), CATSS adapts to examinees' knowledge and adjusts the frequency band of target words accordingly; Laufer & Goldetsin (2004) use *active* and *passive* to refer to what is traditionally considered *productive* and *receptive*
The Productive Vocabulary Levels Test (PVLT)	Productive Form-recall	Laufer & Nation (1995, 1999)	Target is a defining sentence context with a blank for examinees to complete; uses same words and frequency bands as *VLT*	Essentially a form-recall version of the form-recognition *VLT*; though advertised as productive, examinees do not produce items in context of realistic spoken or written discourse (Schmitt, 2010)

Lexical Frequency Profile (LFP), a frequency-based method	Productive Frequency-based	uses *VocabProfile* software developed by Nation, Heatley, & Coxhead	Analyzes lexical output and breaks language into four categories (1K band, 1-2K band, words in AWL, and off-list words)	Analyzes words that appear in lexical output; considered a more *comprehensive* (vs. *selective*) method (Read, 2000); frequency bands based on *General Service List*
BNC 20,000 Profile	Productive Frequency-based	based on work of Nation and Cobb; uses updated version of *VocabProfile*	Analyzes lexical output and reports frequency per every 1,000 words in 1K-20K of *BNC*	Considered a more fine-grained frequency analysis than *LFP* (above); does not contain academic language category; generates lists of inputted words by token, type, and family; available on Tom Cobb's website, *Compleat Lexical Tutor* (http://www.lextutor.ca/tests)
P_Lex	Productive Frequency-based	Meara (2008)	Analyzes lexical output by dividing text into 10-word segments and graphing number of 2K+ frequency words per segment	Single lambda value represents frequency profile; works well with small samples of texts; runs directly from Meara's website (http://www.lognostics.co.uk/tools/index.htm)
V_Size	Productive Frequency-based	Meara & Miralpeix (2008)	Analyzes lexical output and creates lexical frequency profiles (500, 1K, 1.5K, 2K, and 2+K) based on *Zipf's Law*	Offers vocabulary size norms based on theoretical Zipf curves (Miralpeix, 2008)
Lexical diversity (LD)	Productive	N/A	Measures variation in number of word types produced by a learner relative to total number of tokens higher LD reflects richer, more diverse lexicon	Standardized TTRs needed to control for influence of text-length; *Vocd* is available online in CHILDES database software; other LD scores available through a variety of text tools, including *Coh-Metrix*, *TAALES*, and *WordSmith*
TAALES	Productive	Kyle & Crossley (2015)	Analyzes lexical output and generates a number of indices related to frequency and other word properties	Many indices operationalize construct of lexical sophistication; can also be considered a measure of depth of knowledge; *TAALES* can be freely downloaded from www.kristopherkyle.com

Unlike the *VLT*, the *Vocabulary Size Test (VST)* (Nation, 2008; Nation & Gu, 2007) was designed to provide an actual estimate of learners' vocabulary size. The format of the *VST* is a multiple-choice meaning-recognition task in which the target word is presented in a sample sentence as the stem. Target words represent 14 different frequency bands based on the spoken British National Corpus (BNC). In a validity study of the *VST*, Beglar (2010) found that scores on the *VST* decreased significantly toward lower frequency bands, though some high frequency words were surprisingly difficult for learners and vice versa. Dóczi and Kormos (2016) interpret these results to suggest that word properties beyond mere frequency may prove useful in designing test of vocabulary size, and that test developers should taken into account learners' differential exposure to input. Because of the fact that *VST* is a test of overall vocabulary size, Schmitt (2010) suggests that it should have utility in investigations of learners' development progress.

Another popular approach to measuring breadth of knowledge has been the use of a basic checklist, sometimes called a *Yes/No Test* (e.g., Meara, 1992). This type of measure is relatively easy both for test designers and examinees. Learners are simply asked to determine whether they know a lexical item or not; they are not required to produce or recognize its meaning. Such a format is the least time-intensive for test takers and can thus include the largest sample of test items (Schmitt, 2010). Of course, one limitation of the method is that learners aren't actually required to produce word meanings and may over-estimate their knowledge. For this reason, many checklist tests feature the inclusion of non-words so as to identify learners who over-estimate their knowledge or check words indiscriminately (see Huibregtse, Admiraal, & Meara, 2002 for a discussion of potential nonword adjustment methods; Mochida & Harrington, 2006).

The test instruments reviewed above are *selective*, meaning that the items are selected in advance by test designers (Read, 2004). Let us turn to a few examples of breadth of vocabulary measures that would be considered *comprehensive*, meaning that they measure the entire content of examinees' output in a writing or speaking task.

The most widely used comprehensive measure of lexical knowledge in learner output is frequency. Frequency measures used to assess the complexity (Skehan, 2009) or sophistication (Kyle & Crossley, 2015; Read, 2000) of the words learners produce are typically derived from large reference corpora which serve as a proxy for the statistical distribution of words in everyday language use. One of the more popular frequency-based comprehensive measures in vocabulary research is the *Lexical Frequency Profile (LFP)*. Based on the *VocabProfile* software initially designed by Nation, Heatley, and Coxhead (2002), *LFP* analyzes learners' lexical output according to four categories extracted from *General Service List* (*GSL*) (West & West, 1953) frequencies: the first most frequent 1000 words, the second most frequent 1000 words, words in the *Academic Word List* (*AWL*) (Coxhead, 2000), and remaining words that don't fit into any of the previous three categories. A study by Laufer and Nation (1995) found that a cross-section of three proficiency groups demonstrated a stair-step pattern across frequency bands.

While software like LFP use band-based frequency measures, natural language processing tools like *Tool for the Automatic Analysis of Lexical Sophistication* (*TAALES*) (Kyle & Crossley, 2015; Kyle, Crossley, & Berger, 2018) provide count-based indices of frequency in lexical output. *TAALES* produces a variety of indices that measure linguistic features known to index breadth of vocabulary, including frequency and range (i.e., occurrence across genres or text types). Additionally, *TAALES* calculates frequency and range measures for 2-word and 3-word multi-world units (bigrams and trigrams). In a recent validation study, *TAALES* indices were demonstrated to explain over 47% of the variance in written lexical proficiency scores and over 48% of the variance in scores of speaking proficiency. Interestingly, *TAALES* Ngram frequency measures were highly predictive of human ratings of proficiency, both in written and spoken language, while frequency indices based on individual words were not.[2] In a study that examined the relative contribution of both band-based and count-based frequency measures to predicting lexical proficiency, Crossley, Cobb, and McNamara (2013) found that count-based indices were the strongest predictors of lexical proficiency in L2 learners' freewriting. Interestingly, their study also found that beginning learners displayed a greater range of frequency, a finding that would be masked by results based on frequency bands alone. Table 2.1 outlines several other measures that use the frequency-based approach to measuring vocabulary size.

A final approach to breadth of vocabulary involves looking at lexical diversity (LD), which is roughly a calculation of the range of different words a learner produces, under the assumption that learners with larger vocabularies will use a greater variety of words in their repertoire and repeat themselves less. The most basic manner of calculating LD is to observe type-token ratios (TTRs) in learner production; however, because TTR can be impacted by text-length (McCarthy & Jarvis, 2013), more sophisticated LD measures (e.g., Jarvis, 2002; Malvern, Richards, Chipere, & Durán, 2004) have been developed. Like frequency-based comprehensive measures, LD offers no information regarding how well words are used or whether they are repeated for rhetorical or cohesive purposes (Jarvis, 2002; Meara, 2009), nor does it take into account the production of multi-word units (Schmitt, 2010). Nevertheless, Crossley, Salsbury, McNamara, and Jarvis (2010) found that LD (measured with *D*, Malvern et al., 2004) predicted

[2] Note that many of the indices that proved predictive in Kyle and Crossley (2015) measure properties of lexical output that arguably reflect depth of vocabulary knowledge rather than breadth.

34% of variance in holistic lexical proficiency ratings in a corpus of scored L2 writing samples. However, because LD captures repetition, it may also be an indicator of lexical overlap (i.e., cohesion; Crossley & McNamara, 2012), and thus is difficult to interpret.

Breadth of vocabulary estimates are complicated by the fact that researchers do not necessarily agree on what exactly should be counted when measuring vocabulary size, either in creating targets for test items or when measuring learner production (tokens, types, word forms, lemmas, word families, etc.) (Schmitt, 2010). Vermeer (2001) has criticized size-based vocabulary tests for this reason, adding that results across studies will never be comparable so long as researchers fail to agree on what is meant by a "word." Another concern with vocabulary size pertains to the determination of a word's basic meaning and what role partial knowledge may play when assessing breadth of knowledge. Schmitt et al. (2001a) consider this one of the most interesting research directions in vocabulary right now. Another limitation with many of the instruments reviewed above is that they do not offer insights into an examinees' ability to use target words productively in actual communication (Schmitt, Schmitt, & Clapham, 2001). Even the more comprehensive methods (e.g., *LFP*) only analyze words that appear in output, not whether or not they are used well. Finally, all of the instruments reviewed above are specifically designed to measure breadth of English vocabulary. In fact, very few existing vocabulary instruments have multilingual applications.

2.3 Depth of Knowledge

The simplest definition of depth of vocabulary knowledge is "how well particular words are known" (Read, 2004, p. 211; Nation, 2001; Read, 1993). This aspect of knowledge is often attributed to the quality of lexical knowledge, rather than to its size. While we know a great deal about how the form-meaning component of word knowledge is acquired, we know much less about the development of depth of vocabulary knowledge, particularly over time (Dóczi & Kormos, 2016; Schmitt, 2010).

Schmitt (2010) outlines two predominant approaches to conceptualizing depth of knowledge. The first, the "word-centered" approach, is primarily concerned with how well learners know individual lexical items (e.g., Anderson & Freebody, 1981), while the second approach, the "lexicon-based" perspective, is concerned with learners' knowledge of lexical items' relationship to one another in the mental lexicon (e.g., Henriksen, 1999).

Read (2000) initially took an entirely word-centered approach to depth of knowledge when he described *developmentally-focused* vs. *dimensions-based* approaches, neither of which acknowledged the role of lexical organization or automaticity in depth of vocabulary knowledge. The *developmental* approach was concerned with incremental acquisition of lexical items, while the *dimensions-based* approach sought to determine the various components of word knowledge one might have about a particular lexical item. This latter approach is also referred to as a *components* approach (Schmitt, 2010). Since, Read (2004) has revised his taxonomy of approaches to the application of L2 depth of knowledge to include network knowledge. He now proposes three distinct approaches to L2 depth of knowledge investigation: *precision of meaning*, *comprehensive word knowledge*, and *network knowledge* (pp. 211-212). To the extent that is possible, the various approaches to the conceptualization of depth of knowledge have been organized in relation to one another in Table 2.2.

Table 2.2 Approaches to Conceptualizing Depth of Knowledge

Conceptualization	**Sub-categorization**	**Description / definition**
Word-centered (Anderson & Freebody, 1981)	Comprehensive word knowledge (Read, 2004); dimensions/ components approach (Read, 2000; Schmitt, 2010; Nation, 2001)	Type of information learners need to acquire about a given word

	Precision of meaning (Read, 2004); developmental approach (Schmitt, 2010; Read, 2000); partial-to-precise knowledge (Henriksen, 1999)	Describes gradients of word knowledge from zero knowledge to complete mastery
Lexicon-based / network knowledge (Henriksen, 1999; Meara, 1996; Read, 2004)	N/A	Strength, density, and organization of networks linking lexical items in mental lexicon

The discussion of depth of knowledge below is organized according to Read's (2004) three approaches (listed above, also in Table 2.2). Though the three overlap, Read argues that they ultimately differ in how researchers' conceive of depth of knowledge and the type of assessment procedure chosen by researchers to carry out research.

2.3.1 Precision of meaning

Depth of knowledge as conceived through *precision of meaning* often attempts to capture gradations of "elaboration or richness of meaning" (Read, 2014, p. 216). In one regard, this approach overlaps with breadth of knowledge while simultaneously addressing one of its limitations, namely, the inability of breadth of knowledge instruments to capture partial lexical knowledge (Henriksen, 1999; Schmitt et al., 2001). The *precision of meaning* approach to depth of knowledge also bears the most in common with Henriksen's proposed (1999) *partial-to-precise* dimension of lexical knowledge, which assumes that lexical competence equates with precise comprehension. According to Henriksen, "lexical development can be characterized as a move or progression from rough categorization or vagueness to more precision and mastery of finer shades of meaning" (p. 311). The process is often characterized as beginning with word recognition (i.e., recognizing that a word exists in the L2), and then moving through stages from partial to more precise comprehension.

Instruments used to measure *precision of meaning* often employ test items that require precise and specific knowledge, self-report of degrees of knowledge, and elicitation of definitions (Read, 2004), though Henriksen (1999) suggests that translation tasks (L2 to L1), multiple choice definition selection tasks, and paraphrasing tasks can also provide insight into learners' knowledge of precision of meaning. Existing test instruments that measure precision of meaning, and depth of knowledge in general, are compiled in Table 2.3.

2.3.2 Comprehensive word knowledge

Comprehensive word knowledge (as defined by Read, 2004) relates directly to the *components* approach described by Schmitt (2010) and the *dimensions* approach initially outlined by Read (2000) (see Table 3). *Comprehensiveness word knowledge* encompasses aspects of word knowledge that go beyond basic meaning and mostly closely aligns with definitions of depth of meaning that refer to "quality" of knowledge (Read, 1993, p. 357). Nation's (2001) approach to depth of knowledge is typically considered one of the most influential in this respect. Nation referred to vocabulary knowledge as consisting of receptive or productive mastery of information about a word's form, meaning, and use.

The *comprehensive word knowledge* approach to depth of knowledge also overlaps with Henriksen's (1999) second dimension of lexical competence, which she refers to as *depth of knowledge*. The underlying assumption behind Henriksen's depth of knowledge dimension is the complexity of lexical knowledge and "the many types of knowledge that comprise full understanding or a rich meaning representation of a word" (p. 305; see also N. C. Ellis, 1995; R. Ellis, 1995; Harley, 2008; Nation, 1990). Such multidimensional knowledge extends beyond a word's referential meaning to include its sense relations to other words, including both

Table 2.3 Instruments to Measure Depth of Knowledge

Instrument	Approach to Depth	Mastery + Format	Reference	Description	Comments
Vocabulary Knowledge Scale (VKS)	Precision of meaning	Receptive and productive Self-reporting and recall	Paribakht & Wesche (1997); Wesche & Paribakht (1996)	Each item includes a five-level Elicitation Scale combining self-reporting (Levels I-II) with demonstration of knowledge through use (Levels III-V)	Designed to capture initial stages of word learning; intervals between scales are likely not equidistant, preventing parametric statistical analysis (Schmitt, 2010)
Schmitt and Zimmerman scale	Precision of meaning	Receptive and productive "Can-do" self-reporting	Schmitt & Zimmerman (2002); Scarcella & Zimmerman (1998)	Each item includes four self-report options, with the last two asking examinees to report what they can achieve with a word in spoken or written communication	Adapted from *Test of Academic Lexicon* (Scarcella & Zimmerman, 1998); descriptions are more transparent / less metalinguisic than *VKS* (above)
Test of English Derivatives (TED)	Comprehensive Word Knowledge	Productive Form recall	Schmitt & Zimmerman (2002)	Each target word followed by four sentences with blanks; examinees write in the appropriate derivative form (e.g., adjective, noun) of the target lexical item, with the context of the sentences eliciting each derivative	Does not require metalinguistic labels for word class derivations; all target items from 2,000-word *General Service List* (West, 1953)
Discrimination Collocations Test (DISCO)	Comprehensive Word Knowledge	Receptive Form Recognition	Eyckmans (2009)	Two options in each multiple choice item are idiomatic collocations from same frequency band while a third is a free word combination (i.e., not idiomatic); subjects asked to select the two options that are idiomatic and ignore the distractor.	Only measure verb+noun collocations

Word Associates Format (WAF)	Network knowledge, though arguably comprehensive word knowledge as well	Receptive Comprehension of word association	Read (1993, 1998, 2000)	Eight options within two boxes displayed for each target adjective; examinees choose four words that are associates of the target word; all associates in Box 1 are paradigmatically related to the target, while associates in Box 2 are syntagmatically related	Very popular as a depth test; because collocations are included, *WAF* may tap into knowledge of formulaic language as well (Schmitt, 2010)
Lex30	Network knowledge, though arguably comprehensive word knowledge as well	Production Free word association production task	Meara & Fitzpatrick (2000)	Examinees presented with a list of stimulus words and asked to produce at least three single-word associates; stimuli are high frequency and selected to avoid eliciting a predominant response	Meara and Fitzpatrick (2000) analyzed responses according to frequency, while D&K have analyzed responses for concreteness, hypernymy, and frequency (D&K)
TAALES	Comprehensive word knowledge	Free production	Kyle & Crossley (2015)	Analyzes lexical output and generates a number of lexical sophistication indices, including concreteness, imageability, hypernymy, etc.	Many indices operationalize construct of lexical sophistication; also produces frequenty indices for breadth knowledge; TAALES can be freely downloaded from www.kristopherkyle.com

paradigmatic and syntagmatic relations, as well as the syntactic and morphological features of a word (Cronbach, 1942). Here we see echoes of Nation (2001) as well.

A variety of methods have been used to explore depth of knowledge with reference to *comprehensive word knowledge,* including free production tasks, both spoken and written (Crossley, Salsbury, & McNamara, 2010; Crossley et al., 2011; Bell, 2009); interviews (Schmitt, 1996; Schmitt et al., 2001); and word lists, definitions, or gap-fill exercises (Milton & Fitzpatrick, 2014; Zareva, 2005). Some researchers have also employed word association tests (e.g., Fitzpatrick, 2012; Meara, 1983; Read, 1998; Zareva, 2005) in investigation of a dimensions approach to depth of knowledge. However, following Read (2004), network knowledge has been treated in this chapter as a distinct third approach to depth of knowledge. As such, word association tests are discussed further below. Another construct frequently explored in *comprehensive word knowledge* or *dimensions-based* approaches to depth of knowledge is collocational knowledge. The degree to which collocational (or associative knowledge in general) pertains to *comprehensive word knowledge* or *network knowledge* depends on whether the knowledge construct is conceptualized as an aspect of individual word knowledge (cf. Revier, 2009) or as a holistic construct in and of itself and highlights the need to distinguish between the two approaches.

2.3.3 Network knowledge

The final approach, *network knowledge,* conceives of the development of depth of knowledge as the gradual building of a lexical network. Read (2004) explains that

> [a]s a learner's vocabulary size increases, newly acquired words need to be accommodated within a network of already known words, and some restructuring of the network may be needed… This means that depth can be understood in terms of learners'

> developing ability to distinguish semantically related words and, more generally, their knowledge of the various ways in which individual words are linked to each other. (p. 219)

Network knowledge is distinct from *precision of meaning* and *comprehensive word knowledge* to the degree that the focus of analysis is the development of network links between sets of words in the mental lexicon, whereas the first two approaches are primarily concerned with the acquisition of individual words.

In Henriksen's (1999) take on lexical competence, she considered *network building* to be an aspect of depth of knowledge separate from *partial-to-precise knowledge*. Essentially, Henriksen synthesizes what Read (2004) distinguishes (i.e., *comprehensive word knowledge* with *network knowledge*) (see Table 2.2) by suggesting that the following two interrelated processes are equally involved in depth of knowledge: a) the labeling and packaging (Aitchison, 1994) of words as their (primarily extensional) meanings are mapped onto phonological and orthographical forms, and b) "reordering or changing the lexical store via a process of network building" (Henriksen, 1999, p. 309). Henriksen uses the term *semantization* to refer to the process of developing a semantic understanding of a word while simultaneously figuring out its relationship to other lexical items in the mental lexicon. She calls for vocabulary research to be more explicit about which process is being assessed or discussed in investigations of lexical competence and suggests there has been an overemphasis on the first process and a neglect of the second.

2.3.3.1 Word associations

Among the most basic research tools utilized in investigations of network knowledge is the word association (WA) task. Work with learners and L2 WAs first began with Meara and associates (Meara, 1984), with later work by Meara and Fitzpatrick (2000), Schmitt (1998), and Singleton (1999). Meara's early WA studies (and later WA format tests, e.g., Read, 1998) were intended to capture the quality of a learner's lexical competence and the "intensional aspects of a learner's meaning representation" (Henriksen, 1999, p. 306). Initial findings suggested that L2 adult learners tended to produce unstable and irregular responses and that L2 learners, like L1 children, tend to develop from producing collocational associations to more meaning-based ones (Aitchison, 1994; Meara, 1980, 1983).

A few years later, Read (1993, 1998) created a WA instrument for assessing depth of vocabulary knowledge that involved examinees selecting responses to stimuli rather than producing them. Examines had 6-8 potential associates to choose from for each target word. The selected associates were then analyzed according to their potential relationships. Meanwhile, *Lex30* was designed by Meara and Fitapatrick (2000) to measure productive knowledge related to WAs. In *Lex30*, examinees are presented with a list of high frequency stimulus words and asked to produce at least three single-word responses, with stimuli designed so as to elicit a range of potential responses. In a validation study, the frequency of L2 subjects' responses showed a significant positive relationship with subjects' scores on a yes/no checklist, suggesting a relationship between productive depth of knowledge and passive breadth of knowledge (Meara & Fitzpatrick, 2000).

Historically, analysis of L2 WAs have followed one of two approaches. The first approach categorizes the type of relationship between the cue and the association produced, often relying on conventional category distinctions such as paradigmatic (e.g., *apple—fruit*), syntagmatic (e.g., *far—away*), or clang (meaning that the two words are formally related orthographically and/or phonetically; e.g., *height—hike*). The second approach compares L2 learners' WAs to native-speakers' to determine the "canonicity" of L2 responses under the assumption that subjects who produce more native-like associations are more lexically proficient. Such an approach typically involves the use of large-scale, previously collected native-speaker norm lists for reference, such as the Postman-Keppel list (Postman & Keppel, 1970).

Overall, both analytic approaches reveal differences in L1 and L2 WA behavior (Wolter, 2002; Henriksen, 2008; Meara, 2009), but these differences are often "inconsistent and inconclusive" (Fitzpatrick et al., 2013, p. 2), leading some researchers (e.g., Kruse, Pankhurst, & Smith, 1987) to question the value of the word association task as a measure of lexical proficiency entirely. Fitzpatrick and others (Fitzpatrick, 2006; Fitzpatrick et al., 2013) maintain that the potential of the WA task is often masked by problematic practices and assumptions both during data collection and analysis. For example, Fitzpatrick notes the pervasiveness of unprincipled selection of stimulus words, while Fitzpatrick et al. argue against the use of large, outdated native speaker norms to determine canonicity. Instead, they propose the use of so-called "bespoke" norm lists collected from participants who match the demographic profile of the population being assessed instead.

2.3.3.2 Semantic priming

Another source of evidence of emerging L2 lexical network strength can be found in semantic priming (Meyer & Schvaneveldt, 1971; Neely, 1977). Rather than observing learners' ability to produce words that are related to stimulus words, semantic priming focuses on implicit processing of language in a manner that reveals meaning-based links between lexical items. The assumption underlying semantic priming research is that the processing of a stimulus is improved (or otherwise impacted) when it is preceded by a prime with conceptual or associative relations to the stimulus (McNamara, 2005). One of the most widely accepted explanations for semantic priming effects is found in Anderson's model of spreading activation (Anderson, 1983a, 1983b), whereby the activation of a particular lexical item emanates to its neighbors relative to the strength of association between them (Collins & Loftus, 1975). A more strongly associated lexical neighbor is "closer" within the network and thus more likely to be activated than a more distantly related neighbor or a lexical item that is unrelated to the stimulus entirely (Neely, 1991). In simpler terms, semantic priming effects reveal the cognitive mechanisms that "keep tabs" on conceptual and associative elements of the language we encounter (McDonough & Trofimovich, 2009, p. 59). Traditionally, lexical decision tasks have been used to reveal semantic priming, with effects measurable to the degree that researchers observe decreased word recognition latencies between a related prime and target when compared to unrelated pairs (Neely, 1977).

Because lexical decision semantic priming may "provide a window into the development and organization of a lexicon" (McDonough & Trofimovich, 2009, p. 65), it has also been used by researchers interested in the lexical knowledge of L2 learners. Overwhelmingly, L2 priming studies have been conducted to investigate cross-language priming and the degree to which bilingual speakers have shared or separately stored conceptual stores (see Basnight-Brown & Altarriba, 2007 for a review). Less L2 research has been conducted to examine the type of associations that prime lexical processing and the strength of interconnected semantic networks within a single language. Research has indicated, however, that less proficient L2 learners demonstrate stronger semantic priming effects in their L1 than their L2 (Frenck & Pynte, 1987; Phillips, Segalowitz, O'Brien, & Yamasaki, 2004), and that while highly proficient L2 speakers have demonstrated semantic priming effects comparable to NSs, low proficiency L2 learners may show little-to-no priming effects at all (Frenck-Mestre & Prince, 1997). Specifically, studies have found that advanced L2 speakers show facilitatory effects for processing semantic relations in strongly associated synonyms, antonyms, and collocates (Frenck-Mestre & Prince, 1997), but not in weakly associated words (Devitto & Burgess, 2004) or in hyponyms (Crossley, 2013).

An extensive literature review revealed no study that has investigated growth in the development of L2 lexical networks by examining L2 semantic priming effects longitudinally. Nor has any researchers used semantic priming to investigate which types of relationships between words develop first in the L2 mental lexicon. This despite the fact that monolingual priming research has demonstrated that the construct of semantic-similarity is not monolithic and that semantic priming studies make a misstep in treating semantic similarity as "all or none categorical knowledge" (Huettig, Quinlan, McDonald, & Altmann's, 2006, p. 78).

2.3.3.3 Computational measures

Another approach to assessing L2 lexical knowledge—one that could be adapted to investigate network knowledge specifically—involves the computational analysis of learner-produced data. This method typically involves eliciting spoken or written discourse. Analysis of written or transcribed data is then facilitated by natural language processing (NLP) (Meurers, 2013) tools that report numerical indices related to lexical features in the text. While research involving elicited L2 production has been criticized in the past for assuming the representativeness of single isolated utterances produced by learners at one point in time, datasets involving language produced by learners repeatedly over time ensure that acquisition—or lack thereof—is more likely to be evident in data (Norris & Ortega, 2009). Furthermore, when longitudinal corpora are available, it may also be possible to draw conclusions about the development of lexical knowledge over time.

Lexical features investigated in learner output using NLP measures have typically included psycholinguistic word properties, such as frequency, familiarity, age of acquisition, hypernymy, imageability, concreteness, and meaningfulness (Altarriba, Bauer, & Benvenuto, 1999; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Cruse, 1986; A. W. Ellis & Morrison, 1998; Paivio, 1991). NLP approaches have demonstrated strong relationships between computational indices and human ratings of lexical proficiency or at predicting the development of lexical proficiency over time. By and large, the majority of work investigating lexical proficiency based on lexical output has focused on measuring the frequency of the individual words produced (N. C. Ellis, 2002a, 2002b). Other NLP studies based on L2 output have shown that more lexically proficient L2 learners produce less concrete words, less specific words, a greater variety of words, less frequent words, less familiar words, and words with more senses (Crossley & McNamara, 2013; Crossley, Salsbury, & McNamara, 2009, 2010; Crossley, Salsbury, & McNamara, 2011, 2013; Crossley, Salsbury, McNamara, & Jarvis, 2010; Crossley, Salsbury, McNamara, & Jarvis, 2011; Kyle & Crossley, 2015).

While computational indices have proven successful at explaining human ratings of lexical proficiency and lexical growth of time, NLP approaches are often limited to the analysis of the linguistic and psycholinguistic features of individual words. This approach does little to shed light on the manner in which those lexical items are understood in relation to other items in the mental lexicon. For example, Henriksen (1999) has suggested that the type of information about a word that is the most important in terms of lexical network development may depend on word class. Verb-argument patterns are an important meaning representation in knowledge of verbs (Römer, O'Donnell, & Ellis, 2014), while hypernymic relations are important to nouns (Miller & Fellbaum, 1991). One solution is to develop computational indices that measure and assess the relationship of individual words to other words in the lexicon. For example, in addition to existing hypernymy measures (e.g., Crossley, Salsbury, McNamara, et al., 2010), some researchers have begun to use NLP indices to investigate the production of collocations and multiword units (e.g., Crossley & Salsbury, 2011), the variability of semantic contexts in which a word occurs (e.g., Berger, Crossley, & Kyle, 2017b), or the association between lexical items and the more schematic lexicogrammatical constructions in which they frequently occur (e.g., Kyle, 2016).

A limitation of depth of knowledge measures in general is that it can still be quite difficult to determine just how much learners know about given lexical items. Furthermore, when just a few words are assessed for depth of knowledge—as is often the case—it can be difficult to generalize vocabulary knowledge (Read, 2004). Read (2004) cautions researchers against using the term "depth of knowledge" indiscriminately and suggests that any measure designed to assess vocabulary knowledge should be explicit in defining what is measured. Another limitation of most depth of knowledge measures is that they are designed to assess learners' static lexical knowledge at one point in time rather than their lexical development over time, despite the increasing evidence that words are learned incrementally (Dóczi & Kormos, 2016; Read, 2004). More longitudinal studies are also needed in order to adequately describe lexical development in network building (Henriksen, 1999).

2.4 Longitudinal Development of Lexical Knowledge

2.4.1 Longitudinal development of breadth of knowledge

As mentioned above, the majority of instruments designed to measure breadth of knowledge do so by analyzing lexical frequency (either of target items or of lexical output). This is largely due to the assumption, attested by evidence from language acquisition studies from a range of paradigms, that more frequent lexical items are acquired earlier than lexical items that are less frequent in input (N. C. Ellis, 2002a, though see Beglar, 2010 for an exception). For this reason, the validation tests mentioned above often evaluated how well test scores or items discriminated across groups of learners at different pre-determined proficiency levels, with varying educational experience, etc., in hopes that these cross-sectional datasets might serve as a proxy for the manner in which language learners actually develop lexical knowledge over time. In doing so, Dóczi and Kormos (2016) argue that these cross-sectional studies “describe a state of learners’ knowledge rather than a path of development” (p. 29). For researchers interested in observing patterns of growth as breadth of vocabulary knowledge develops over time, more longitudinal studies of breadth of vocabulary are needed.

On the whole, longitudinal development remains an underexplored phenomenon in SLA (Ortega & Byrnes, 2008). Furthermore, despite recent claims that language learning is a non-linear process (e.g., Larsen-Freeman, 2009), many language studies have no choice but to restrict themselves to linear models because they have so few data collection points over time (Ortega & Byrnes, 2008; Ortega & Iberri-Shea, 2005). According to Ortega and Iberri-Shea (2005), a truly longitudinal study can be identified by its length, multiple waves of data collection, a study design focused on capturing change, and the investigation of antecedent-consequent relationships rather than cause-and-effect (i.e., change in context vs. as a result of experimental manipulation). While relatively more longitudinal research has been conducted to examine the development of vocabulary size than depth of vocabulary knowledge, the lack of longitudinal work in breadth of vocabulary studies still presents a serious research gap in the field (M. Daller, Turlik, & Weir, 2013; Dóczi & Kormos, 2016).

One exception is a longitudinal study by Zhang and Lu (2013) that analyzed the development of vocabulary size using the *VST* (Nation, 1983, 1990; Schmitt et al., 2001) to investigate Chinese university students' English L2 development at three different points in time over 22 months. Results demonstrated a significant effect of frequency level for both breadth of knowledge and fluency development. Specifically, learners demonstrated a 35% rate of improvement for the 3K frequency list and *AWL*, and a 100% improvement on words in the 5K frequency band. While the pattern of development for the 5K frequency band was linear, learning at all other bands followed the power law of learning (Newell & Rosenbloom, 1981), which maintains that "the rate of improvement is reduced as practice continues" (Fitts & Posner, 1967, p. 18).

Meara (1992) predicted this flattening out years ago, expecting a plateau in frequency development as learners acquired words beyond the 5000-word frequency level. Such a plateau is also predicted by researchers who assume that the frequency with which words occur in input impacts their acquisition (N. C. Ellis, 2002b; Read, 1998; N. Schmitt, N. Schmitt, & C. Clapham, 2001b) or that there is a systematic inverse relationship between a word's rank frequency and its actual frequency of occurrence (Zipf, 1935, 1949). The latter would mean that the most frequent word in a language occurs approximately twice as often in input as the second most frequent word, resulting in learning that is relatively steep at first, followed by a flattening out as less frequent words are encountered less frequently in input.

While the majority of empirical studies that support such a prediction (see Milton, 2009 for a review), Dóczi and Kormos (2016) note that all of them analyzed cross-sectional data rather than longitudinal. To address this gap, Dóczi and Kormos describe a study in which they administered sections of the *VST* (Nation & Beglar, 2007) and the *PVLT* (Laufer & Nation,

1999) to Hungarian L1 EFL students and university-level ESL students in Britain from a variety of L1 backgrounds. Learners took the two tests at the beginning of their study and at the end of an 8-month interval. Both groups made significant improvements on both tests during that time. While the EFL learners' rate of improvement on both the *VST* and the *PVLT* was higher at the end of 8 months, Dóczi and Kormos suggested that this may be due to the power law of learning, as the EFL group's initial knowledge was significantly less than the ESL learners. They also hypothesize that the differences in learning rate could be due to differences in learning context. Frequency analysis found that word frequency contributed the most to explaining variance in both tests' pre- and post-scores and to gains over time, with target word occurrence in the CELEX corpus (Baayen, Piepenbrock, & Gulikers, 1995) explaining more variance than frequency bands derived from the BNC. This last finding confirms Crossley, Cobb, and McNamara's (2013) finding regarding the greater predictive strength of count-based frequencies relative to band frequencies.

Dóczi and Kormos' (2016) study investigated the longitudinal development of breadth of vocabulary using primarily selective assessment measures. In a more comprehensive approach, Crossley, Salsbury, and McNamara (2010) analyzed the spoken lexical output of six ESL learners collected every two weeks over the course of an entire year. Using frequency counts based on CELEX word frequency values, the authors found that the most significant frequency growth occurred in the first four months of intensive language study, after which frequency levels of produced content words leveled out. In this regard, the learners in this study exhibited a breadth of vocabulary learning curve based on the power law of learning, "where a steep increase at the beginning is followed by a flattening out of the learning curve in later stages" (M. Daller et al., 2013, p. 213).[3] Another important result of Crossley, Salsbury, and McNamara was that learners actually produced less frequent words at the beginning the study and more frequent ends toward the end, a finding that has been replicated in other studies as well (Crossley, Salsbury, Titak, & McNamara, 2014; Crossley, Skalicky, Kyle, & Vanderbilt, in press).

Finally, in one of the few longitudinal studies investigating rate of acquisition, Schmitt and Meara (1997b) administered the *VLT* (Nation, 1990) to high school and university Japanese L1 EFL learners both prior to and following one year of study. They found that on average, learners gained recognition knowledge of 330 words, though individuals patterns of development showed a great deal of fluctuation, with 28% of learners demonstrating actual decreases in recognition knowledge. Schmitt and Meara's findings suggest that breadth of knowledge develops in a non-linear fashion and may be incremental.

Taken as whole, the existing longitudinal studies that have investigated growth in breadth of vocabulary over time reveal that the frequency of words in L2 input may be a significant driver of vocabulary size, with actual count-based frequencies offering more explanatory power than frequency bands (Crossley, Cobb, & McNamara, 2013; Crossley, Salsbury, & McNamara, 2010; Dóczi & Kormos, 2016). The literature also suggests that L2 vocabulary size develops in an incremental and non-linear fashion and may be impacted by context of learning and the type of L2 input that different contexts provide.

2.4.2 Longitudinal development of depth of knowledge

Of the few longitudinal studies that have investigated depth of knowledge, most have looked at interactions between breadth and depth of vocabulary knowledge over time. For example, Schmitt and Meara (1997a) observed how EFL learners' word associations (WAs) and

---

[3] See Section 2.42 for further discussion of this study from a depth of knowledge perspective.

grammatical suffix knowledge (i.e., depth of knowledge) developed over the course of a year in relation to general language proficiency and breadth of vocabulary knowledge—based on the TOEFL (Educational Testing Service, 1987) and the *VLT* (Nation, 1990), respectively. Scores for the above were obtained both at the beginning and at the end of the year. Results demonstrated that while WA knowledge and knowledge of suffixes (i.e., both depth of knowledge measures) correlated with one another and with scores on the TOEFL and *VLT*, subjects showed no actual improvement in WA or verbal suffix knowledge over the course of the year. However, by the end of the year, subjects displayed 19-25% more receptive breadth of knowledge than productive breadth of knowledge.

In another study investigating the interaction between breadth and depth of lexical knowledge, Zareva (2005) set out to empirically test Henriksen's (1999) three-dimensional global framework for lexical knowledge and determine the effectiveness of each in distinguishing between three proficiency levels of English speakers: NESs, intermediate, and advanced EFL speakers. Zareva conducted a regression analysis to determine how well five variables (self-perception of vocabulary knowledge, knowledge of words from different frequency bands, vocabulary size, number of associations, native-likeness of associations) commonly used to assess lexical competence could predict participants' vocabulary knowledge. The five variables were determined using a modified version of Wesche and Paribakht's (1996; Paribakht & Wesche, 1997) *VKS* test. All five predictors were significant for both NES and NNSs. However, only self-perception of vocabulary knowledge bore a significant relationship to participants' overall vocabulary knowledge. Follow-up analyses demonstrated that combinations of predictors worked differently for different proficiency groups, with the smallest "best" combination a two-predictor model consisting of verifiable self-report and vocabulary size. The full five-predictor model also performed well but was deemed impractical from an assessment standpoint and was not as good as the two-predictor model at explaining variance in lexical knowledge of participants at different proficiency levels. Overall, results suggest that the relative importance of Henriksen's three dimensions may change with proficiency and that receptive-productive control and quantity were more predictive of variation in lexical knowledge than quality/depth (operationalized here as WAs).

Nurweni and Read (1999) found similar results in their year-long study that analyzed the development of breadth and depth of lexical knowledge vocabulary in first year EFL learners at an Indonesian university. Breadth of knowledge was assessed with a word translation task and depth with a WA test. The authors found the strength of the positive relationship between breadth and depth varied depending on proficiency, with a much higher correlation for students with high proficiency, suggesting that breadth and depth may be more likely to converge at more advanced levels of proficiency.

In Crossley, Salsbury, and McNamara's (2010) year-long longitudinal study (reviewed in previous section), the authors found that the frequency of L2 learners' produced content words (i.e., breadth of knowledge) leveled out after four months. However, additional quantitative and qualitative analysis of produced words' meaning senses (i.e., depth of knowledge) demonstrated that learners began to use the extended meanings of polysemous words already in their repertoire following the initial four months of rapid learning. This finding suggests that increases in depth of knowledge may occur after a certain threshold of breadth of knowledge has been achieved and that a plateau in breadth of knowledge does not necessarily indicate a similar plateau in the development of depth of knowledge.

In a longitudinal study examining the spontaneous speech of the same group of learners, Crossley, Salsbury, and McNamara (2009) found that hypernymic relations in learners' lexical output increased over the course the year-long data collection. In other words, learners produced more abstract language over time. An additional finding was a significant correlation between learners' hypernymic language production (i.e., depth of knowledge) and lexical diversity (i.e., breadth of knowledge), suggesting that the two developed in tandem. Taken together, Crossley, Salsbury, and McNamara (2010) and Crossley, Salsbury, and McNamara (2009) suggest that both the mode of lexical production (i.e., written or spoken production) and the particular operationalization of breadth and depth of knowledge my impact longitudinal findings.

In another investigation of the relationship between depth and breadth of vocabulary knowledge, Qian (1999, 2000) attempted to predict scores on the reading section of the TOEFL with vocabulary size estimates obtained from *VLT* (Nation, 1990) and depth of knowledge obtained from a version of Read's (1998) WAs test. Qian found a strong correlation between breadth of knowledge and depth of knowledge, suggesting that "development of the two dimensions is probably interconnected and interdependent…. [and may be due to] partial construct overlap of the two measures" (p. 299). That said, in a regression analysis, each test still made a unique contribution to explaining variance in scores on the reading test. Thus, while breadth and depth develop in tandem, they still may index related but distinct constructs.

Fitzpatrick (2006, reviewed above) compared NNSs' WA responses (i.e., depth of knowledge) to the size of their receptive vocabulary (breadth of knowledge), as determined by their scores on the yes/no *Eurocentres Vocabalary Size Test* (Meara & Jones, 1990). There was no overall difference in NNSs' broad response categories as a result of proficiency, though there was a significant positive correlation between phrasal collocation (e.g., *method* => *madness*) and proficiency and a significant negative correlation between proficiency and similar form association (e.g., *undertaking* => *funeral*). Fitzpatrick concluded that the assumption of an un-native-like to native-like continuum in WA response behavior with proficiency may be problematic. She also suggests that NSSs' responses are more likely constricted by weak or nonexistent links in the lexicon rather than acquisition or availability of individual words.

It is worth mentioning that not all researchers have always taken a dichotomous approach (e.g., Anderson & Freebody, 1981) to vocabulary knowledge. For example, Vermeer (2001) is a strong advocate of a non-dichotomous network-based approach to breadth and depth that makes essentially no conceptual distinction between the two. In a study that examined breadth and depth of vocabulary knowledge in L1 and L2 children learning Dutch, Vermeer found high correlations between two traditional breadth measures and a depth measure (a receptive knowledge, a word description, and a guided association task, respectively), suggesting that the more words a child knows, the more she can describe it in depth. Ultimately, Vermeer takes a network knowledge approach in arguing that if a breadth of knowledge test contains a principled sample of words, it can measure depth as well as breadth: "The denser the network around a word, the richer the set of connections around that word, the greater the number of words known, and the deeper the knowledge of that word" (p. 231).

While Read (2004) warns that Vermeer's (2001) data may have been highly influenced by factors related to the cognitive development of young children, he does question whether any measure of depth of knowledge we currently have can do a better job than a well-designed test of vocabulary size and suggests that breadth and depth of vocabulary knowledge are best understood from a lexical network perspective and likely to develop in parallel. Such an assumption is "particularly pertinent if we adopt a network building perspective on depth, in that vocabulary growth also entails the building of more extensive linkages between items in the mental lexicon” (Read, 2004, p. 221).

2.5 Overview of Lexical Development Research

2.5.1 Challenges and future directions

Before concluding, I now turn to existing challenges and research gaps in the study of L2 lexical knowledge. I include an expanded discussion of the potential for psycholinguistic methods to offer insights into L2 lexical knowledge and how the various aspects of L2 lexical knowledge interact over time.

2.5.1.1 Improved depth of knowledge measures

According to Read (2004), "A broader range of measures is needed before we can be more confident about the extent to which depth in some sense makes a contribution to the assessment of the lexical knowledge of L2 learners” (p. 223). For example, if vocabulary development is truly incremental (e.g., Schmitt & Meara, 1997), then we need more measures that capture partial knowledge of words. The creation of a valid developmental scale for vocabulary has proven quite difficult: It is still unclear how many stages of acquisition there should be, what role receptive and productive knowledge might play in gradation, and what zero- vs. complete-knowledge actually looks like (Dóczi & Kormos, 2016; Read, 2000; Schmitt, 2010). Furthermore, any developmental approach to precision of meaning knowledge assumes that vocabulary knowledge proceeds in a linear and sequential fashion (Dóczi & Kormos, 2016), an assumption that the studies reviewed in breadth of knowledge do not support. Read (2004) also suggests that current approaches to precision of meaning implicitly reinforce the “adult native speaker criterion” made explicit in Anderson and Freebody’s (1981)’s definition of depth (see p. 10 above).

Furthermore, while *network knowledge* approaches to depth of knowledge have relied almost exclusively on word association (WA) data, what L2 researchers actually learn from NNS association data is still a matter of debate (Fitzpatrick, 2006), in part because L2 associations are not predictable (Riegel & Zivian, 1972; Söderman, 1993) and may have more to do with culture than proficiency (Kruse, Pankhurst, & Smith, 1987). Furthermore, it is rarely clear what type of knowledge the studies that incorporate WA data are attempting to operationalize, especially when they fail to specify whether their conceptualization of depth of knowledge is conceived as comprehensive word knowledge (i.e., word-centered) or as the strength of lexical network connections (i.e., a more holistic construct). The use of WA data to draw conclusions about word-centered depth of knowledge is problematic to the extent that L2 association responses may actually be constricted by weak links in their lexicon rather than by the availability of individual words (Fitzpatrick, 2006).

2.5.1.2 Less focus on individual word knowledge

Paradis (2009) has warned that assessing learners' knowledge of single words taps exclusively into learners' explicit, declarative lexical knowledge, regardless of how proficient of a speaker they may be: "Single words fundamentally differ from the rest of language" (Paradis, 2009, p. 184). Evidence from studies investigating words in context, lexical network knowledge, and multi-word units suggest that conclusions drawn from learners' production, comprehension, access, or processing of individual words may lack psychological validity. This has implications not only for the incorporation of collocations, multi-word units, and other phraseological units in studies of L2 lexical knowledge (e.g., Biber & Conrad, 1999; Granger & Meunier, 2008; Siyanova & Schmitt, 2008; Wray, 2002), but also for more approaches to depth of knowledge that examine the size, density, and organization of links in the mental lexicon (e.g., H. Daller et al., 2007).

2.5.1.3 More longitudinal studies

Larsen-Freeman's (2009) mandate to "think longitudinally and nonlinearly" when we investigate L2 acquisition holds for studies of lexical knowledge as well (p. 584). While there is a scarcity of longitudinal studies in SLA as whole (Ortega & Iberri-Shea, 2005), this is especially true for L2 research that investigates lexical development and network building (Dóczi & Kormos, 2016; Henriksen, 1999). Furthermore, to date, little work has been done to empirically explore the relationship between the various components of depth of knowledge as they develop over time (Read, 2004).

With regard to the relationship between development over time and proficiency, Ortega and Byrnes (2008) have recommended that researchers interested in true longitudinal linguistic development may need to operationalize the construct of "advancedness" as distinct from skills or proficiency. Such an approach implies a distinction between the constructs of development and proficiency, one advocated by other researchers of late (e.g., Hulstijn, 2011). For example, Hulstijn (2011) has proposed a distinction between basic language cognition (BLC) and higher language cognition (HLC), arguing that while all NSs have BLC (regardless of literacy, age, etc.), NSs do not necessarily share equivalent HLC. Meanwhile, the notion of native-like attainment for language learners only pertains to BLC of the L2. Attributing a more fundamental role to speech than literacy (e.g., Bloomfield, 1933; De Saussure, 1916), Hulstijn argues that BLC pertains specifically to automatized speech production and reception but does not include reading or writing. HLC is thus an extension of BLC that includes less frequent lexical and morphosyntactic structure, as well as written language.

Any developmental investigation of L2 BLC or HLC would thus require analysis of a longitudinal spoken dataset, something lacking in the majority of studies that have investigated size or quality of L2 lexical knowledge. The over-reliance on written data in L2 lexical studies is especially limiting given that the controlled manner of written production may not capture all aspects of lexical knowledge (see Hulstijn, 2011).

2.5.1.4 More focus on procedural L2 vocabulary knowledge

The distinction between BLC and HLC (Hulstijn, 2011) is tangentially related to Read's (2004) call for more L2 vocabulary assessment measures that target procedural (vs. declarative) knowledge. To the degree that productive mastery of L2 lexical knowledge may demonstrate proceduralized knowledge, Meara's (1997) claim that productively-known lexical items are those that can be activated by their links to other items in the lexical network may prove insightful. In this regard, Meara takes a network approach to the construct of productive knowledge. The assumption here is that words that are only known receptively do not have enough links in the lexicon and thus are not sufficiently activated for recall or use (i.e., production). To my knowledge, no study has empirically teased out the relationship between productive vocabulary mastery, procedural knowledge, and strength of network knowledge as inter-related constructs.

2.5.1.5 Psycholinguistic approaches to investigating l2 lexical knowledge

One potential avenue to developing better methods of measuring L2 procedural vocabulary knowledge is through the use of psycholinguistic approaches. While SLA has borrowed constructs from cognitive psychology for some time, such as procedural vs. declarative use of knowledge (Anderson, 1983a) or automatic processing (Schiffrin & Schneider, 1977), the field has not traditionally borrowed its operationalizations (Sanz & Grey, 2015). Rather, SLA has traditionally over-relied on accuracy measures. According to Sanz and Grey (2015), the field has only very recently incorporated reaction times as a dependent measure, which are “better suited to characterize qualitative changes in processing and knowledge” (p. 317).

While more and more psycholinguistic methods are gradually being used to investigate L2 phenomena, very few involve the analysis of data sets with more than one data collection. Those that have involved multiple psycholinguistic datasets from the same participants (e.g., Akamatsu, 2008) typically collected data immediately prior to and following a treatment or training procedure, rather than conducting multi-wave data collection over a significant amount of time (Ortega & Byrnes, 2008; see Segalowitz, Segalowitz, & Wood, 1998 for an exception). More longitudinal psycholinguistic studies of this nature would allow L2 researchers to examine the growth of breadth and depth of lexical knowledge over time, as it develops in the same learners. RTs are particularly conducive to longitudinal analysis, as it is the case that latencies in isolation do not provide immediately insights into cognition or language processing. In a cross-sectional study, a participant's absolute speed on any given task is not as informative as how rapidly she performs different types of tasks, relative to other participants (Jiang, 2012). In the case of longitudinal analysis, researchers are able to use subjects’ previous RTs as a baseline for additional comparison.

2.5.2 Summary and conclusion

This chapter began by examining the traditional distinction between breadth and depth of L2 lexical knowledge in the field of SLA and reviewing the various ways in which these constructed have been defined and operationalized, with a focus on lexical network approaches to conceptualizing depth of knowledge and the mental lexicon. Next, attention was given to empirical evidence that offered insights into the manner in which these various conceptualizations of breadth and depth of L2 lexical knowledge develop with respect to one another and over time. Finally, I acknowledged existing challenges and future directions in the study of L2 lexical knowledge, many of which emerged in earlier discussions.

Some researchers have begun to question whether the dichotomous breadth vs. depth paradigm is still useful (e.g., Read, 2004). In a way, such researchers are not so much concerned with the distinction between breadth and depth as they are the monolithic conceptualizations of each and the outdated approaches used to measure them. For example, Read (2004) recommends that we prioritize explicit and comparable definitions of the construct meaning(s) operationalized and consider the theoretical basis behind the various assessment measures. Henriksen (1999) makes a similar recommendation, arguing that "if we want to take the first tentative steps in the direction of developing a unified theoretical construct of lexical competence and a model of vocabulary development, it is necessary to strive for more precision and standardization" (p. 304). Ultimately, lexical knowledge remains "many-faceted," with no currently available method or measure capable of tapping into "all forms of lexical knowledge" (Schmitt et al., 2001a, p. 61). It is likely that the best insights into understanding the multi-dimensional nature of L2 lexical knowledge and its development over time will emerge from the accumulation of evidence from a diverse variety of paradigms and methodological approaches.

3 PROJECT OVERVIEW

This dissertation project takes up the challenges posed in Chapter 2 through three related focal studies that track the development of depth of lexical knowledge in adult L2 English language learners. To do so, the project examines the potential for elicited word associations, lexical decision semantic priming, and spoken lexical production to offer insight into the strength and organization of L2 lexical networks as they develop over time. In doing so, the results of this project aim to offer researchers multidimensional, quantitative models of the organization and development of depth of L2 lexical knowledge from a lexicon-based perspective.

Throughout this chapter and those that follow, the use of the terms *network knowledge* and *lexical networks* is intended to distinguish the approach taken here from *precision of meaning* and *comprehensive word knowledge* approaches to depth of knowledge (Read 2004; Schmitt 2010; also see Chapter 2). Specifically, the focus of analyses described below was the development of network links between sets of words in the L2 mental lexicon, whereas the other two approaches are primarily concerned with the acquisition of individual words independent from knowledge of other words.

The purpose of Chapter 3 is to provide a brief overview of this dissertation project. Prior to outlining the three focal studies comprising this project, research subjects who participated in the studies are introduced along with descriptive statistics regarding their linguistic backgrounds and demographic information.

3.1 Participants

Research participants for the studies described in Chapters 4-6 consisted of adult English learners recruited from two Intensive English Programs (IEP) at two universities in the Southeastern United States, as well as undergraduate native English speakers (NES) and proficient functionally bilingual speakers of English enrolled at a state university in the Southeastern United States. Each participant group is described in further detail below. Unless otherwise stated, all L2 learners (i.e., IEP students) took equivalent versions of the focus experiments up to four times (every two months for up to six months), while undergraduate (UG) subjects participated in focal experiments only once.

Prior to data collection, all research subjects completed a demographic questionnaire with information related to gender, age, handedness, and additional language(s) spoken. Any subject who did not list English as a native language(s) answered further questions about their L1 background, L2 onset age, years of L2 classroom learning, and years of residence in L2 environment. The demographic questionnaire can be found in Appendix A.

In addition to the three focal studies, all subjects were also asked to take a vocabulary size test (adapted from the *Vocabulary Levels Test*; Schmitt, Schmitt, & Clapham, 2001). L2 learners completed the vocabulary size test as well as a brief Language Experience Survey during each of the scheduled data collections. The vocabulary size test and Language Experience Survey are described below following additional details regarding participant groups.

3.1.1 L2 Learners

L2 learners were recruited through classroom visits or through emails sent by their IEP instructors. Only those subjects who began studying English in an IEP program in the spring semester of 2017 and placed into high beginner through high intermediate levels in their respective programs were recruited. To incentivize participation, L2 learners were compensated with electronic Amazon gift cards each time they participated in a data collection.

Thirty-four L2 learners were initially recruited for participation in this dissertation project. Of those 34, 21 contributed meaningful data (i.e., the subject completed a demographic questionnaire, Language Experience Survey, vocabulary size test, and at least one of the three focal studies) twice or more during longitudinal data collection. Demographic information for these 21 learners is listed in Table 3.1, while dominant and additional languages spoken by L2 learners are listed in Table 3.2. Note that 25 dominant languages are listed because some participants reported more than one native language. In addition, seven subjects claimed to speak an additional language(s) beyond their native language(s) and English, though their self-reported knowledge of such languages was below the ability to speak and understand "practically any form of the language with nuance and precision."

Table 3.1 L2 Learners' Demographic Information

	N	Age		Gender			Hand			L1 distance	
		Mean	**SD**	**Male**	**Female**	**Other**	**Right**	**Left**	**Both**	**Mean**	**SD**
L2 Subjects	21	28.238	8.372	6	15	0	21	0	0	1.97	0.477

Table 3.2 L2 Learners' Language Backgrounds

Language	Dominant	Additional
Arabic	1	1
Azerbaijani	1	0
Mandarin Chinese	5	0
French	2	1
Lingala	1	0
Japanese	2	1
Persian	1	0
Portguguese	1	0
Spanish	7	3
Turkish	2	0
Vietnamese	1	0
Russian	1	1
Korean	0	1
German	0	1
Catalan	0	1

**More than 21 dominant languages are listed because some participants listed more than one native language*

3.1.2 Undergraduate subjects

NES and functionally bilingual participants were recruited from an undergraduate subject pool or participated for extra credit and thus did not require monetary compensation. All NES and functionally bilingual subjects were enrolled as undergraduates (UGs) at Georgia State University. A total of 79 UG subjects contributed meaningful data to at least one focal experiment in this dissertation project. Demographic information for these 79 subjects is listed in Table 3.3.

Table 3.3 Undergraduates' Demographic Information

	N	Age		Gender			Hand		
		Mean	**SD**	**Male**	**Female**	**Other**	**Right**	**Left**	**Both**
UG	79	23.987	8.413	23	54	2	73	6	0
NES	54	25.130	9.923	17	35	2	48	6	0
FB	25	21.520	1.806	6	19	0	25	0	0

Throughout this project, the term *NES* will be used to refer to subjects who reported English as their only native language. Meanwhile, the term *functionally bilingual* (FB) is used to refer to undergraduate subjects who reported a language other than—or in addition to—English as their dominant language. While such subjects likely exhibit varying degrees of English proficiency, their status as undergraduate students at an American university indicates that they are able to use English for sophisticated purposes across a variety of genres and modalities. Language background information for all UG subjects is reported in Table 3.4.

Table 3.4 Undergraduates' Language Backgrounds (other than English)

Language	NES	Functionally Bilingual (other than English)	
	Additional	**Dominant**	**Additional**
American Sign Language	5	1	0
Bengali	8	2	0
Chinese	0	4	1
French	0	6	0
German	3	1	1

Italian	2	0	2
Japanese	5	2	5
Korean	1	2	2
Portuguese	0	1	3
Somali	0	1	0
Spanish	15	2	4
Thai	0	2	0
Vietnamese	0	1	0

Thirty-nine NES subjects claimed to know an additional language(s) beyond English, though they did not list this language(s) as "native" or "dominant," and their self-reported knowledge of the language(s) was below the ability to speak and understand "practically any form of the language with nuance and precision." Twenty-two FB subjects reported speaking an additional language(s) beyond English and/or their native language(s), though their self-reported knowledge of any such languages was also below the ability to speak and understand "practically any form of the language with nuance and precision."

3.2 Language Experience Survey

In addition to providing demographic information at the time of initial data collection, L2 subjects completed a brief Language Experience Survey (LES) prior to each longitudinal data collection. The LES asked subjects information about their average exposure to and use of English. Given that not all IEP experiences in the United States are consistent, the purpose of the LES was to determine each L2 learners' level of exposure to and usage of English. The survey was loosely adapted from questions developed by Csillagh (2015, 2016) and made available by the IRIS Respository (Marsden, Mackey, & Plonsky, 2016). The LES used during this project can be found in Appendix B.

The LES was programmed in the online survey software Qualtrics and completed by L2 learners each time they participated in a data collection. Learners' responses to the LES were used to derive three language experience measures analyzed as independent variables in the three studies that comprise this dissertation project. The three measures were self-reported proficiency, exposure, and use[4]. See appendix C for further details regarding how these measures were calculated. The results of L2 subjects' three LES measures over time are reported in Table 3.5.

Table 3.5 L2 Learners' Longitudinal Language Experience Survey Results

	N	**Self-reported proficiency**		**Exposure**		**Use (hours of EN per day)**	
		Mean	SD	Mean	SD	Mean	SD
Time 1	21	1.524	0.602	2.439	0.609	3.143	1.621
Time 2	21	1.714	0.644	2.624	0.510	3.048	1.431
Time 3	18	1.889	0.758	2.539	0.482	3.333	1.572
Time 4	14	1.857	0.864	2.486	0.494	3.643	1.447

**Sample size indicates number of subjects who produced usable data per time period*

3.3 Vocabulary Size Tests

A vocabulary test was administered to all L2 participants and 41 UG subjects[5] (29 NES and 12 FB) to determine their vocabulary size in English (i.e., breadth of knowledge) and to compare L2 development that emerged in focal studies with growth in vocabulary size. Unfortunately, to date, no well-validated vocabulary size instrument exists that offers four equivalent versions for longitudinal data collection[6]. Nor was it deemed efficient to give lengthy vocabulary tests to subjects at the beginning of each data collection. For this reason, two existing equivalent versions of the *Vocabulary Levels Test* (*VLT*; Schmitt, Schmitt, & Clapham, 2001) were divided into four shorter versions.

While initially designed as a diagnostic instrument, rather than a size test per se, the *VLT*

[4] Specifically, *use* was calculated based on subjects' response to the following question: "How many hours per day do you usually spend speaking English?" (see Appendix C for details).

[5] Some UG data was collected prior to beginning administration of the VLT; thus, not all UG participants have VLT scores.

[6] An exception is the *Yes/No Test* (e.g., Meara, 1992), but this format was deemed too similar to semantic priming, which is employed in Study 2.

is considered one of the best-known and most widely used standard vocabulary measures in the field (Meara, 1996; Read, 2005; Schmitt, 2010). The original format of the *VLT* was developed by Nation (1983) but later modified by Laufer and Nation (1999), Beglar and Hunt (1999), and N. Schmitt, D. Schmitt, and C. Clapham (2001a). The *VLT* measures learners' basic knowledge of a word's form-meaning relationship at four different frequency levels (the most frequent 2K, 3K, 5K, and 10K words). Both a receptive and a productive version of the *VLT* exist. The receptive format, used here, involves a form-recognition matching task in which test takers are asked to select the appropriate word from six options to match each of three brief definitions (i.e., stems). This form-recognition matching format offers three stems and six options per "cluster," with 10 clusters presented for each of several frequency bands, for a total of 30 items per frequency band per test version.

For the current project, the two existing versions of the receptive *VLT* were divided into four equivalent sub-versions. Sub-versions included five clusters (15 items) from each of the 2K, 3K, 5K, and 10K word levels, for a total of 20 clusters (60 items) per test. Each subtest contained the same number of nouns, verbs, and adjectives (see Table 3.6) and maintained the equivalent ratio of nouns to verbs to adjectives found in Schmitt, Schmitt, and Clapham (2001). The resulting four tests were programmed in the online survey software Qualtrics and randomly assigned to subjects over the course of the study. UG subjects were randomly assigned one of the four VLT sub-versions. The four sub-versions of the VLT used in this dissertation study can be found in Appendix D.

Descriptive statistics for all participants' VLT scores can be found in Table 3.7. Group means for L2 learners' VLT scores (Table 3.7) indicated a potential trend of growth in vocabulary size over time. In order to determine whether this growth was statistical, and whether demographic variables, individual differences, or subtest version impacted vocabulary size, a linear mixed effects models (LME) was conducted predicting VLT scores.

Table 3.6 Vocabulary Subtests Derived from VLT (Schmitt, Schmitt, & Clapham, 2001), Versions 1 and 2

Part of speech (POS)	POS per subtest	VLT Subtest 1				VLT Subtest 2				VLT Subtest 3				VLT Subtest 4			
		2K	3K	5K	10K	2K	3K	5K	10K	2K	3K	5K	10K	2K	3K	5K	10K
Nouns	10	2	3	2	3	3	2	3	2	2	3	2	3	3	2	3	2
Verbs	6	2	1	2	1	1	2	1	2	2	1	2	1	1	2	1	2
Adjectives	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
TOTAL	20	20				20				20				20			

Table 3.7 VLT Mean Scores, Descriptive Statistics

Participant Group		**VLT score (60 possible)**	
	N	Mean	SD
L2 Learners			
Time1	21	30.682	12.461
Time 2	21	32.905	10.672
Time 3	18	33.722	13.974
Time 4	14	35.643	14.232
Undergraduates	41	56.561	4.879
Native English Speakers (NES)	29	58.207	1.780
Functionally Bilingual (FB)	12	52.583	7.354

For this model, age, gender, L1 distance[7], VLT score, self-reported proficiency, exposure (mean score based on survey results), average hours of English used per day (i.e., use), VLT subtest version, and time were included as fixed effects. For the categorical variables, the following baselines were set: *gender*: male; subtest*: subtest 1;* and *time*: Time 1. Subjects were entered as random effects. This model found no significant differences in L2 learners' VLT scores over time, nor were there significant differences across subversions of the VLT. There were no significant main effects for demographic variables or individual differences. The model reported a marginal R^2 of 0.125 and a conditional R^2 of 0.686. Post-hoc analysis found no significant differences for time or subversion when variables were re-leveled for comparison. Visual inspection suggested the model was not impacted by heteroscedasticity (in other words, residuals were evenly distributed). Table 3.8 displays the coefficients, standard error, *t* values, and *p* values for the L2 learner model predicting VLT score.

[7] L1 distance ranges from 1 to 3 (with a score of 3 reflecting a language very close to English) and is calculated based on the difficulty of learning L2 English as a function of one's L1 (Chiswick & Miller, 2005).

Table 3.8 Linear Mixed Effects Model Predicting VLT Scores Among L2 Learners

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)	9.429	15.176	0.621	0.542
Demographic Variables				
Age	0.327	0.328	0.995	0.334
Gender (baseline: male)	7.311	5.810	1.258	0.227
Individual Differences				
Hours per day	-0.737	0.986	-0.747	0.458
L1 distance	3.226	5.226	0.617	0.547
Self-proficiency	-0.477	2.399	-0.199	0.843
Exposure	1.641	3.078	0.533	0.596
Procedural Variables				
Sublist (baseline: Subtest 1)				
Subtest 2	-3.188	2.685	-1.187	0.242
Subtest 3	2.189	2.589	0.845	0.403
Subtest 4	-1.019	2.591	-0.393	0.696
Experimental Variables				
Time (baseline: Time 1)				
Time 2	2.696	2.540	1.061	0.294
Time 3	2.040	2.631	0.775	0.442
Time 4	4.400	2.899	1.518	0.136

**Significant predictor of VLT score (p<0.05)*

For the UG model, a linear regression[8] was conducted predicting VLT using age, gender, and language status (NES or FB[9]) as dependent variables. Results found no significant differences in VLT score across subversions of the test. However, there were significant differences between NES and FB undergraduates, with NES subjects scoring significantly higher on the VLT than FB subjects (see Table 3.7). There were no other significant main effects and post-hoc analysis found no significant differences for VLT subversion when this variable was re-

[8] A linear mixed effects model could not be conducted on the UG VLT data because there were no repeated measures.

[9] Note that while 79 UG participants contributed meaningful data to this dissertation project only 41 completed the VLT (29 UG and 12 FB).

leveled for comparison. Table 3.9 displays the coefficients, standard error, *t* values, and *p* values for the UG model predicting VLT score. The model reported an adjusted R^2 of 0.369.

Table 3.9 Linear Regression Model Predicting VLT Scores Among Undergraduates

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**60.071**	**2.929**	**20.510**	**<.001**
Demographic Variables				
Age	0.052	0.081	0.644	0.524
Gender (baseline: male)				
female	-1.657	1.604	-1.033	0.309
other	1.383	3.447	0.401	0.691
Individual Differences				
Status (baseline: NES)*	**-6.081**	**1.561**	**-3.896**	**0.000**
Procedural Variables				
Subtest (baseline: Subtest 1)				
Subtest 2	-1.681	1.977	-0.850	0.401
Subtest 3	-3.172	2.006	-1.581	0.123
Subtest 4	-2.917	2.020	-1.444	0.158

**Significant predictor of VLT score (p<0.05)*

3.4 Focal Studies

3.4.1 Study 1: Word associations

A common tool used to investigate both psychological and word knowledge information is the free word association (WA) task. WA tasks involve presenting a word (i.e., a cue) to subjects and asking them to speak or write the first word(s) that comes to mind. While word associations have been often used in the field of psychology, they are less commonly conducted in subjects' L2. Even more rare are longitudinal studies that track how L2 word associations change over time as subjects gain exposure to the language being studied.

Chapter 4 describes a study (Study 1) in which WA data was collected longitudinally from L2 learners and from UG subjects cross-sectionally. Data were collected digitally through an online survey. Follow-up questions during the experiment asked subjects to indicate their level of familiarity with the cue word they had seen previously. L2 learners' WAs were compared to a preexisting set of NES WA norms, as well as to the associations produced by UG subjects in response to the same set of cues. Results were further analyzed in order to observe changes in the native-likeness of L2 learners' lexicons with relation to time and vocabulary size.

3.4.2 Study 2: Semantic priming

Lexical decision semantic priming involves asking subjects to read one word and then determine whether a subsequent string of letters is a real word or not. Evidence of priming manifests as faster response times when a real word (e.g., *nurse*) has been briefly preceded by a semantically related word (e.g., *doctor*) than by an unrelated one (e.g., *spider*) (Neely, 1977), with the assumption that faster processing is due to stronger links in the networks between the two related words. The accumulation of evidence from L2 semantic priming studies suggests that facilitation effects of semantic priming are much stronger for proficient L2 subjects than less proficient ones (Frenck-Mestre & Prince, 1997; Vasos, 1983); however, to date, attempts to investigate L2 semantic priming longitudinally have not occurred. Furthermore, some studies have found no priming effects for weakly associated words (Devitto & Burgess, 2004) or hyponyms (Crossley, 2013) even in advanced L2 speakers. Nor is it clear which types of associations among words develop first in L2 learners.

The goal of the study (Study 2) described in Chapter 5 was to observe the effects of L2 semantic priming in L2 learners over time so as to track the development of their associational knowledge in English. Like Study 1, Study 2 also sought to characterize the organization of L2 learners' lexicons and compare them to the lexicons of UG subjects. Stimuli for Study 2 were designed so that the influence of the type of relationship between cues and targets (e.g., Is the prime a synonym of the target or an antonym?) on observed priming effects could be observed and traced over time.

Study 2 data were collected digitally using the online experimental software, Testable. In each trial, the prime word briefly appeared on screen and was automatically replaced by either the target word or a nonsense word (e.g., *blerk*). Subjects were instructed to respond by pressing one key if the target word was indeed a real word in English or a different key if it is a nonsense word in English. In this regard, Study 2 incorporated a lexical decision task. Results were analyzed in order to determine whether evidence of priming in L2 learners increased over time and what relationship the prime-target relationships (e.g., synonyms vs. common collocates) had to time and vocabulary size. UG priming data were analyzed as well for the sake of baseline comparison.

3.4.3 Study 3: Lexical output

A final approach to assessing L2 lexical knowledge involves the computational analysis of learner-produced data using natural language processing (NLP) (Meurers, 2013) tools to obtain numerical indices related to lexical features in the text. Recently, computational indices have been developed to measure the relationship of individual words to other words in the lexicon and to assess meaning in language that extends beyond the single word in isolation (Cowie, 1998; Granger & Meunier, 2008; Granger & Paquot, 2008).

Study 4 used videos to elicit spoken narrative re-tellings. L2 learners watched a brief, 3-minute nonverbal video and were then instructed to describe what they had just seen in three minutes or less. This speech was transcribed and analyzed using automated computational measures selected to assess the associative properties of lexical items produced as learners gain exposure to English. Learner speech was also compared to their own vocabulary size scores.

3.5 Conclusion

The goal of this chapter was to provide an overview of this dissertation project and the three focal studies that comprise it. Focal studies were designed to assess the longitudinal development of adult L2 English learners' depth of lexical knowledge over seven months of language study in the United States. The studies employ diverse experimental methods, including English word associations (Study 1), lexical decision semantic priming (Study 2), and the computational analysis of subjects' spoken output (Study 3) to assess lexical network knowledge. These same methods were applied to undergraduates' lexical network knowledge to obtain a baseline model of how English lexical knowledge is organized. In addition to the above data, all subjects completed a demographic questionnaire, Language Experience Survey, and a vocabulary size test.

Results from this project have the potential to make important contributions to existing assessment research by investigating patterns of development in L2 language English lexical knowledge in a manner heretofore under-examined due to a lack of longitudinal datasets. Findings may also advance theoretical understandings in the fields of SLA and assessment by offering models of productive and receptive English lexical network knowledge and of how such knowledge develops in L2 English learners over time. In addition, results can contribute knowledge to testing specialists and curricula designers who wish to better understand the lexical properties of the English language and predict the development of lexical knowledge with relation to time or general vocabulary size. Finally, in offering one of the first descriptions of how English L2 lexical network knowledge develops, the results of this project could enable assessment stakeholders to establish empirically grounded measures for lexical knowledge and realistic benchmarks for L2 learning.

4 WORD ASSOCIATIONS: STUDY 1

A tool used to investigate both psychological and word knowledge information is the free word association (WA) task. WA tasks involve presenting a word (i.e., a cue) to research subjects and asking them to speak or write the first word(s) that comes to mind. The assumption behind such a task is that WA behavior reveals the patterns of connections between items in the mental lexicon (Henriksen, 1999, 2008). While WAs have been used a great deal in the field of psychology, they are less commonly conducted with bilingual subjects or L2 learners.

Analyses of existing L2 WAs have found that responses are not entirely predictable (Riegel & Zivian, 1972; Sökmen, 1993) and that it is still difficult to determine what is learned from WA data (Fitzpatrick, 2006, 2012b; Fitzpatrick et al., 2013) due to a number of reasons. First, a great deal of WA research is flawed by methodological limitations, such as exclusive use of high frequency stimuli words (e.g., Namei, 2004; Orita, 2002); failure to control for part-of-speech of cues (see Bagger Nissen & Henriksen, 2006); reliance on a small number of cues (e.g., Grabois, 1999; van Ginkel & van der Linden, 1996); classification of learners' responses into a few broadly defined categories (Greidanus & Nienhuis, 2001; Politzer, 1978; Söderman, 1993); the analysis of learner response words according to association type or frequency alone (e.g., Meara & Fitzpatrick, 2000; Namei, 2004); and comparison of L2 word associations to large native-speaker norms that do not match the target population (see Fitzpatrick et al., 2013; Schmitt, 1998). Moreover, the majority of L2 WA studies are based on cross-sectional datasets. Few, if any, studies have investigated the longitudinal development of word association data over time (known exceptions are Dóczi & Kormos, 2016; Fitzpatrick, 2012a; M. Randall, 1980; Schmitt & Meara, 1997) or its relationship to lexical network knowledge, specifically.

The goal of Study 1 (S1) was to investigate the development of English learners' lexical network knowledge through the lens of L2 learners' longitudinal WA behavior. Another goal is to determine which manner of analyzing WAs (see below) offers the best model of L2 lexical development over time. This chapter is motivated by the following questions:

1a. How do L2 learners' English WAs develop with relation to time or vocabulary size?

2a. Which manner of analyzing the canonicity of L2 WAs—comparison to a large native-speaker norms list or a bespoke dataset—bears the strongest relationship to time or vocabulary size?

4.1 Method

The online survey platform, Qualtrics was used to collect written word association (WA) data from L2 learners, as well as from undergraduate (UG) subjects. The experiment presented subjects with a single real word in English (i.e., a *cue*) and asked them to type the first English word that came to mind. Because subjects produced a single response to each word, Study 1's experiment can be described as a discrete association task. Such a task was selected for two reasons: 1) L2 subjects' responses could then be compared to a large pre-existing dataset of discrete associations (see below), and 2) recent studies (e.g., Nelson, McEvoy, & Dennis, 2000) have demonstrated that the first response provided in word associations tasks is typically a more reliable indicator of a word's strongest associates.

Upon completion of a pre-determined number of word associations (30 for L2 learners and 60 for UGs), subjects were presented again with the previously seen words and asked how familiar they were with the word (with "3" indicating a word that they could confidently use in a sentence, "2" indicating a word that they are somewhat familiar with, etc.). The specific experimental procedure for S1 is explained in more detail below following an description of S1 participants and stimuli.

4.1.1 Participants

S1 analyzed the WA responses of L2 learners (*n* = 20) longitudinally, and of UG (*n* = 37; 26 NES and 11 FB), subjects cross-sectionally. UG participants contributed to WA data one time, while L2 subjects were asked to complete a similar WA experiment once every two months for seven months (four times total). Chapter 3 offers further details regarding subjects' recruitment and demographic and language backgrounds.

4.1.2 Stimuli

In total, 120 English cue words were selected as stimuli for the word association task. From these 120 words, four subsets of cue words (30 each) were randomly assigned to L2 subjects across the four data collection periods. UG subjects, who only participated in S1 one time, were randomly assigned two of the subsets (60 cues each) (see Table 4.1). Cue words were balanced across the four subsets for frequency and part of speech (POS), and were controlled statistically for length, concreteness, and number of commonly associated word types in a previously existing NES WS dataset. Cues were also selected to ensure that they had not typically elicited a single, dominant response or idiosyncratic responses in NES data.

Each subset of 30 English cues in Study 1 contained 10 words from each of the following three frequency bands based on the most frequent 5,000 lemmas in the Contemporary Corpus of American English (COCA): 1000-word, 2000-3000 word, and 4000-5000 word bands (Table 4.2). POS was also balanced among frequency bands and among subsets, with each subset containing five nouns, three verbs, and two adjectives per frequency band (15 nouns, 9 verbs, and 6 adjectives total per subset) (Table 4.1).

Concreteness values were derived from Brysbaert, Warriner, and Kuperman (2014). In order to determine a cue word's common associations among NESs, existing large-scale NES WA data were analyzed (cf. Wolter, 2002) in the Edinburgh Associative Thesaurus (EAT, Kiss et al., 1973). Among other things, the EAT reports the number of word types produced in response to a given word when it appeared as a cue in a word association task with native English speakers.

Table 4.1 Study 1 Stimuli: Distribution of Part of Speech per Experiment Subset

Subsets		0-1K Frequency			2-3K Frequency			4-5K Frequency			Total Cue Words	
		Nouns	Verbs	Adjs	Nouns	Verbs	Adjs	Nouns	Verbs	Adjs	UG	L2
UG A	L2 A	5	3	2	5	3	2	5	3	2	60	30
	L2 B	5	3	2	5	3	2	5	3	2		30
UG B	L2 C	5	3	2	5	3	2	5	3	2	60	30
	L2 D	5	3	2	5	3	2	5	3	2		30
Totals		30			30			30			120	

Note. Word class distribution was designed to roughly approximate distribution of word class in English, following Schmitt, Schmitt, & Clapham, 2001

Table 4.2 lists descriptive statistics for cue words' length, concreteness, and number of common association types. No significant differences were found among stimuli in subsets for word length, $F(3, 116) = 0.064$, $p = 0.979$); concreteness, $F(3, 116) = 0.048$, $p = 0.986$); or number of association types, $F(3, 116) = 0.399$, $p = 0.754$). Finally, none of the cue words used in Study 1 occurred as primes or targets in any of the four semantic priming stimuli subsets found in Study 2 (see Chapter 5).

Table 4.2 Study 1 Stimuli Descriptives for Word Length, Concreteness, and Number of NES Associations

	Subset	N	Mean	Std. Dev.	95% C.I. for Mean	
Length	1	30	5.467	1.479	4.914	6.019
	2	30	5.467	1.502	4.906	6.028
	3	30	5.567	1.431	5.032	6.101
	4	30	5.600	1.545	5.023	6.177
	Total	120	5.525	1.472	5.259	5.791
Concreteness	1	30	3.331	1.020	2.950	3.712
	2	30	3.259	1.185	2.817	3.702
	3	30	3.358	1.041	2.969	3.747
	4	30	3.342	1.126	2.922	3.763
	Total	120	3.323	1.082	3.127	3.518
Common word associations (types)	1	30	41.400	10.411	37.513	45.287
	2	30	41.433	11.473	37.149	45.717
	3	30	41.533	10.884	37.469	45.598
	4	30	43.900	9.553	40.333	47.467
	Total	120	42.067	10.523	40.165	43.969

** Number of word association types derived from Kiss et al. (1973)*
*** Concreteness scores derived from Brysbaert et al. (2013)*

4.1.3 Procedure

Upon beginning the online experiment for S1, subjects were first presented with a single real word in English. They were given eight seconds to read this cue word and type the first English word that came to mind. To illustrate, Figure 4.1 shows a sample screen in which a subject saw the word “broken” and typed “window.”

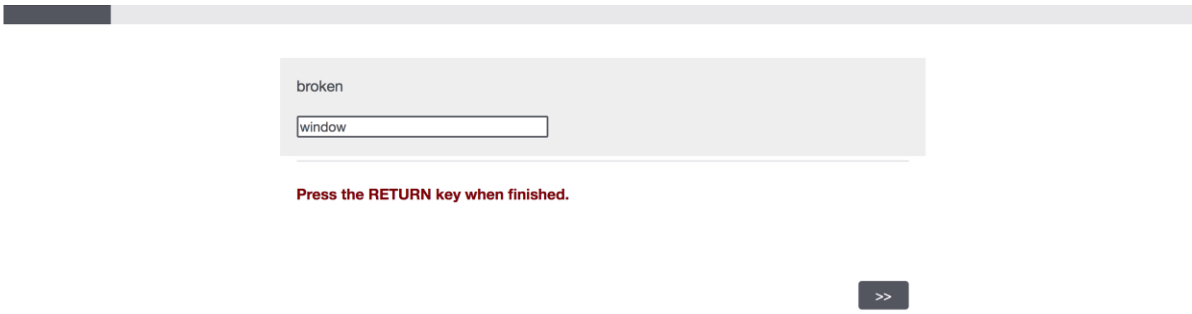

Figure 4.1 A sample of the initial screen seen by subjects during the word association task.

To avoid self-priming, trials were separated by simple math problems (for an example, see Figure 4.2). Subjects were required to answer each math problem correctly before advancing to the next screen. The purpose of the math problem was to distract subjects so that the words appearing in previous trials did not prime their associations to new cue words.

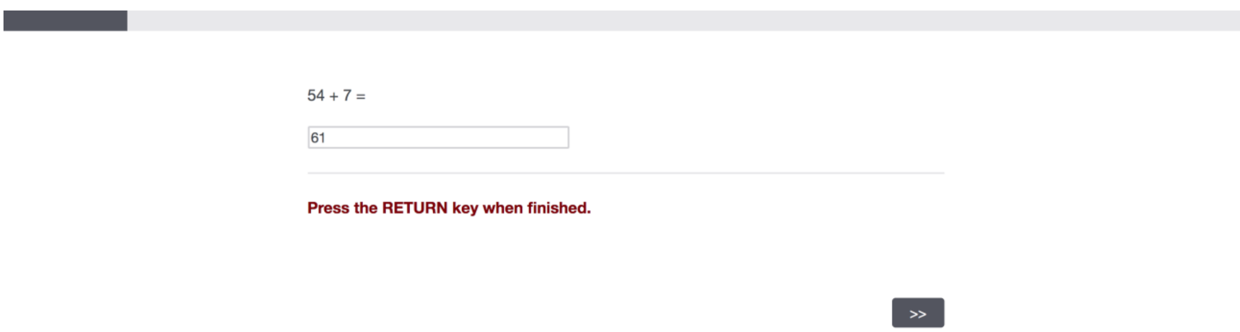

Figure 4.2 A sample mathematical problem separating word association trials.

Upon completion of the WA portion of the experiment (L2 subjects responded to 30 cues per experiment, while UG subjects responded to 60; see breakdown below), subjects were told that they would encounter their responses again. Subjects were presented with a previously seen cue word and asked to select the multiple-choice option that indicated how familiar they were with the word. Figure 4.3 shows a sample screen in which subjects were asked to select their level of familiarity with the cue word “broken.”

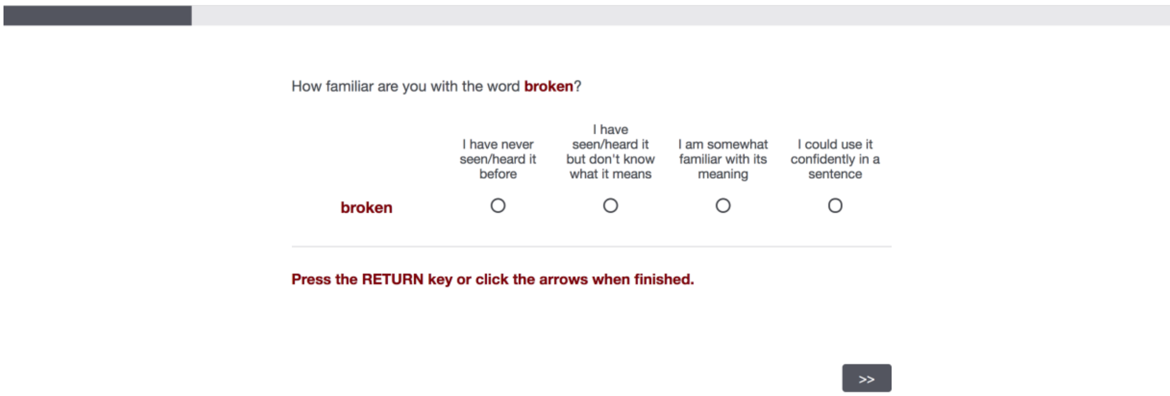

Figure 4.3 A sample screen asking how familiar subjects are with the cue word "broken."

4.1.4 Analysis

The above data were analyzed in order to characterize the nature of L2 learners’ lexicons as compared to undergraduates’ and to observe changes over time in L2 learners’ lexicons with relation to time and vocabulary size. Specifically, canonicity analysis (i.e., how stereotypical subjects’ responses are in relation to a comparison group) was used to determine the “native-likeness” of subjects’ responses.

In the current study, canonicity entailed the comparison of L2 learners’ responses to both pre-existing NES word norms and to UG responses collected in the current study. First, existing University of South Florida (USF) word norms (Nelson et al., 2004) were analyzed. USF is a dataset set of word association norms collected from NES undergraduates in the United States in 1973. The dataset contains empirical association norms for 5,019 English cue words. In addition to comparing L2 learners’ responses to USF word norms, L2 WA responses were compared to so-called “bespoke” WAs collected from UG subjects during the course of the study. Bespoke UG norms were analyzed in order to compensate for the likelihood of generational differences between the current study's research participants and USF participants (see Fitzpatrick, Playfoot, Wray, & Wright, 2013 and Schmitt, 1998 for further rationale underlying collection of bespoke norms list and canonicity analysis).

In total, 2010 L2 WA observations and 2430 UG WA observations were collected. Prior to analysis of L2 WAs, any response which was left blank or for which a subject responded "I don't know" (or some variant thereof) was removed. In addition, WAs for which subjects had indicated "I have never seen/heard it [the cue] before" or "I have seen/heard it [the cue] but don't know what it means" in response to the cue word were removed as well. This was done in order to ensure that WA behavior (i.e., depth of knowledge) was analyzed rather than mere breadth of knowledge (i.e., form-meaning pairings). After removal of WAs based on failure to respond or no/little self-reported familiarity, a total of 1528 observations remained for analysis. These WAs were manually reviewed and it was determined that all respondents had taken the task seriously (e.g., no subject produced nonsense strings of letters or simply repeated the same word over and over again).

Next, following criteria developed by Zareva and Wolter (2012), inflectional (and certain derivational) variants of words were lemmatized. Subjects' lemmatized WAs were then compared to pre-existing USF and bespoke norms lists to determine their canonicity. A subject's response was considered canonical with regard to the USF if it was produced by at least two or more subjects in the USF dataset. A "1" was given to responses that met this criterion and "0" to any response that did not. A response was considered canonical with regard to the bespoke dataset if it was produced by at least one undergraduate research subject in the current study. Again, a "1" was assigned to responses that met this criterion and "0" to any response that did not. In this manner, each L2 learner received a USF canonicity score and a separate bespoke canonicity score per cue word, per data collection (see Hirsh & Tree, 2001; Kruse et al., 1987; Meara, 1978; and Miller & Chapman, 1983 for previous examples of a similar methodology).

In order to determine how L2 learners' WAs developed longitudinally, mixed effects logistic regression analysis was used to predict the canonicity of subjects' WA responses. Separate models for USF canonicity and bespoke canonicity were constructed. For the both models, time, gender, age, L1 distance[10], VLT score, self-reported proficiency, exposure (mean score based on survey results), use (average hours of English used per day), and familiarity were included as fixed effects. For the categorical variables, the following baselines were set: *gender*: female; *handedness*: left; *time*: Time1; *relatedness*: unrelated. An interaction between time and VLT was also investigated. Subjects and cues words were entered as random effects.

4.2 Results and discussion

4.2.1 Descriptive statistics

Table 4.3 reports descriptive statistics for Study 3, including mean and standard deviations for subjects' average USF and bespoke canonicity scores, with L2 learners' scores separated by time of data collection.

Table 4.3 Descriptive Statistics for Word Association Canonicity Analysis

	USF Canonicity		**Bespoke Canonicity**	
L2 Learners	**Mean**	**SD**	**Mean**	**SD**
Time 1	0.517	0.137	0.421	0.130
Time 2	0.479	0.139	0.401	0.140
Time 3	0.453	0.122	0.375	0.122
Time 4	0.444	0.150	0.359	0.134
UG learners	0.567	0.115	0.549	0.133
NES UGs	0.589	0.119	0.563	0.145

[10] L1 distance ranges from 1 to 3 (with a score of 3 reflecting a language very close to English) and is calculated based on the difficulty of learning L2 English as a function of one's L1 (Chiswick & Miller, 2005).

FB UGs	0.516	0.092	0.514	0.098

4.2.2 USF Canonicity Model

An initial model specified with an interaction between time and VLT failed to converge, so a new model was refit with no interactions. The resulting linear mixed effects logistic regression model is reported in Table 4.4, which displays the coefficients, standard error, *t* values, and *p* values for the USF model.

Table 4.4 Study 1 Mixed Effects Logistic Regression Model Predicting Canonicity of Word Associations Based on USF

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**-2.512**	**0.973**	**-2.581**	**0.010**
Demographic Variables				
Age	0.009	0.014	0.692	0.489
Gender: male	0.019	0.260	0.075	0.941
Individual Differences				
Hours per day	-0.047	0.061	-0.767	0.443
L1 distance	0.083	0.210	0.396	0.692
Self-proficiency	-0.037	0.130	-0.281	0.779
Exposure	0.257	0.192	1.340	0.180
VLT	0.006	0.008	0.780	0.435
Familiarity*	**0.441**	**0.186**	**2.371**	**0.018**
Procedural Variables				
Sublist (baseline: Sublist 1)				
Sublist 2	-0.314	0.306	-1.027	0.305
Sublist 3	-0.111	0.321	-0.347	0.729
Sublist 4	-0.469	0.311	-1.509	0.131
Experimental Variables				
Time (baseline: Time 1)				
Time 2	-0.365	0.188	-1.939	0.053
Time 3*	**-0.394**	**0.186**	**-2.118**	**0.034**
Time 4	-0.347	0.197	-1.766	0.077

**Significant predictor of canonicity (p<0.05)*

Significant ($p < 0.05$) and approaching significant ($p < 0.10$) variables from the model in Table 4.4 were then entered into the final linear mixed effects logistic regression model reported in Table 4.5. This final model predicting USF canonicity for L2 learners found significant differences in the canonicity of WAs produced over time, with Time 1 scores demonstrating significantly greater canonicity than Time 3. Post-hoc contrast analyses demonstrated that there were no other significant differences in canonicity between time periods. Familiarity was another significant main effect, with learners producing more canonical associations in response to words with which they were more familiar.

The model in Table 4.5 reported a marginal R^2 of 0.007 and a conditional R^2 of 0.275. Visual inspection suggested the model was not impacted by heteroscedasticity. Table 4.5 also displays the coefficients, standard error, *t* values, and *p* values for the L2 learner model.

Table 4.5 Study 1 Linear Mixed Effects Logistic Regression Model Predicting Canonicity of Word Associations Based on USF, Significant Variables Only

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**-1.599**	**0.667**	**-2.397**	**0.017**
Individual Differences				
Familiarity*	**0.403**	**0.176**	**2.295**	**0.022**
Experimental Variables				
Time (baseline: Time 1)				
Time 2	-0.221	0.170	-1.299	0.194
Time 3*	**-0.359**	**0.175**	**-2.046**	**0.041**
Time 4	-0.303	0.185	-1.634	0.102

**Significant predictor of canonicity (p<0.05)*

4.2.3 Canonicity Model

Because a bespoke model specified with an interaction between time and VLT failed to converge, a new model was refit with no interactions. The resulting linear mixed effects logistic regression model is reported in Table 4.6, which displays the coefficients, standard error, *t* values, and *p* values for the bespoke model.

Table 4.6 Study 1 Linear Mixed Effects Regression Model Predicting Canonicity Based on UG Bespoke Data

Fixed Effect		Coefficient	Std. error	*t*	*p*
(Intercept)*		**-2.540**	**0.957**	**-2.655**	**0.008**
Demographic Variables					
Age		-0.005	0.013	-0.364	0.716
Gender: male		-0.137	0.246	-0.559	0.576
Individual Differences					
Hours per day		-0.029	0.060	-0.474	0.635
L1 distance		0.007	0.194	0.037	0.970
Self-proficiency		0.035	0.126	0.278	0.781
Exposure		0.320	0.190	1.682	0.092
VLT		0.013	0.007	1.716	0.086
Familiarity*		**0.426**	**0.189**	**2.249**	**0.025**
Procedural Variables					
Sublist (baseline: Sublist 1)					
	Sublist 2*	**-0.610**	**0.301**	**-2.024**	**0.043**
	Sublist 3	-0.055	0.314	-0.176	0.860
	Sublist 4*	**-0.778**	**0.308**	**-2.524**	**0.012**
Experimental Variables					
Time (baseline: Time 1)					
	Time 2	-0.366	0.191	-1.920	0.055
	Time 3*	**-0.373**	**0.188**	**-1.986**	**0.047**
	Time 4*	**-0.466**	**0.198**	**-2.354**	**0.019**

**Significant predictor of canonicity (p<0.05)*

Significant ($p < 0.05$) and approaching significant ($p < 0.10$) variables from the model in Table 4.6 were then re-entered into models until only significant variables remained. This final model predicting bespoke canonicity for L2 learners (Table 4.7) found no significant ($p < 0.5$) difference in the bespoke canonicity of WAs produced over time, though a trend was evident (*p*

< 0.10) whereby learners produced less canonical word associations at Time 3 and Time 4 than at Time 1. Post-hoc contrast analyses confirmed this trend ($p < 0.10$). Sublist and familiarity were also significant main effects, with learners producing more canonical responses to words with which they were more familiar.

The model in Table 4.7 reported a marginal R^2 of 0.029 and a conditional R^2 of 0.256. Visual inspection suggested the model was not impacted by heteroscedasticity. Table 4.7 also displays the coefficients, standard error, *t* values, and *p* values for the L2 model.

Table 4.7 Study 1 Linear Mixed Effects Logisic Regression Model Predicting Canonicity Based on UG Bespoke Data, Significant Variables Only

Fixed Effect		Coefficient	Std. error	*t*	*p*
(Intercept)*		**-1.620**	**0.691**	**-2.346**	**0.019**
Individual Differences					
Familiarity*		**0.406**	**0.180**	**2.259**	**0.024**
Procedural Variables					
Sublist (baseline: Sublist 1)					
	Sublist 2	-0.551	0.302	-1.826	0.068
	Sublist 3	-0.002	0.310	-0.005	0.996
	Sublist 4*	**-0.723**	**0.304**	**-2.378**	**0.017**
Experimental Variables					
Time (baseline: Time 1)					
	Time 2	-0.197	0.174	-1.133	0.257
	Time 3	-0.328	0.178	-1.837	0.066
	Time 4	-0.364	0.186	-1.955	0.051

**Significant predictor of canonicity (p<0.05)*

4.2.4 Post-hoc Analysis

Because it was possible that the inclusion of FB undergraduates in the calculation of the bespoke canonicity scores may have influenced results, a post-hoc analysis was conducted predicting a bespoke canonicity measure derived from NES undergraduates' WAs only. Results from this model demonstrated a similar (non-significant) trend toward less bespoke canonicity over time among L2 learners and produced the same significant main effects found in Table 4.7.

4.2.5 Discussion

This study used linear mixed effects logistic regression analyses to determine how L2 English learners' canonical word associations (WAs) develop longitudinally over six months. Pre-existing NES norms (specifically, the USF word association dataset) were used to develop one measure of canonicity of subjects' WAs, while a second dataset of WAs derived from UGs who also contributed to the current study was analyzed to obtain a second measure of bespoke canonicity. In addition to time, other variables included in the mixed effects analysis included demographic information (e.g., age), individual differences (e.g., L1 distance), and procedural variables (e.g., stimuli sublist).

Descriptive analysis of canonicity scores revealed a trend whereby L2 learners produced less canonical WAs over time (see Table 4.3). This trend was evident in both USF and bespoke canonicity scores. Linear mixed effects logistic regression analysis of USF scores demonstrated significant differences ($p < 0.05$) in the canonicity of L2 WAs between Time 1 and Time 3 (Table 4.5), with learners producing less canonical WAs at Time 3 than at Time 1. Analysis of the bespoke corpus demonstrated differences approaching significance ($p < 0.10$) between Time 1 and Time 3, and between Time 1 and Time 4 (Table 4.7). Again, learners produced less canonical responses over time. These differences held even when demographic variables, individual differences, and procedural variables were controlled in analysis.

Given that findings were similar between the main bespoke analysis, which included FB and NES participants, and the post-hoc analysis, which only included NES—and because the purpose of gathering a bespoke dataset was to compare subjects' WAs to their peers' (which includes functionally bilingual English speakers at the university)—the remainder of this discussion will refer to the original analysis based on bespoke canonicity derived from both NES and functionally bilingual undergraduates.

These findings can be interpreted in a number of ways. One explanation for the decrease in canonicity is procedural: Even though data collection only took place once every two months, it may be that L2 learners tired of participating in the WA task and thus began to produce responses that were more random and less relevant to the cue word in later months. Another interpretation calls into question the very assumption that L2 learners would necessarily produce more native-like WAs as they develop lexical proficiency. For example, it may be that learners' responses grow more lexically sophisticated (Kyle & Crossley, 2015), though less canonical, with time. It could also be the case that L2 development does not converge on NES norms at earlier stages of development, as has been demonstrated with frequency in longitudinal L2 lexical production (Crossley et al., in press). Indeed, Fitzpatrick (2006) has suggested that the notion of an un-native-like to native-like continuum in WA response behavior with proficiency may be problematic, while others have argued that differences in responses between native and non-native speakers could have more to do with culture than with proficiency (Kruse, Pankhurst, & Sharwood Smith, 1987).

A third interpretation of these findings is with reference to what did not predict the canonicity of WAs; in particular, vocabulary size had no impact on the native-likeness of learners' WAs. Recall that 181 observations for which subjects had little-to-know familiarity with the cue word were purposefully removed from analysis so that lack of knowledge of specific words could not influence results; thus, the inclusion of vocabulary size as a dependent variable represents breadth of lexical knowledge overall rather than knowledge of particular words that appeared as stimuli in the current study. With this in mind, it could be argued that the lack of predictability of vocabulary size on WAs is because WA behavior is, in fact, less a result of the availability of individual lexical items in L2 learners' repertoire than of the strength of *links* in learners' lexicons, as suggested by Fitzpatrick (2006). Because L2 learners' vocabulary size showed no significant change over time (see Chapter 3 and Tables 3.7 and 3.8) but canonicity did, it seems likely that the WA task and the VLT are indeed measuring distinct (albeit overlapping) constructs.

To better understand—and thus interpret—the trend toward gradually less canonical WAs observed in these data, post-hoc qualitative analysis of four subjects who demonstrated decreases in one or both methods of measuring canonicity was conducted. These included L2 subjects 2, 7, 22, and 31. Demographic information for these four learners can be found in Table 4.8, along with their canonicity scores at each time period. The framework used to assess the relationship between cues and responses produced by learners was drawn from a combination of previous taxonomies (Albrechtsen, Haastrup, & Henriksen, 2008; Fitzpatrick, 2006; Fitzpatrick & Izura, 2011; Fitzpatrick et al., 2013). This categorization framework, comprising six broad categories (including "Other") with 18 subcategories in total, is summarized in Appendix E.

Table 4.8 Qualitative Analysis of Four L2 Learners' Word Associations

Qualitative analysis of four L2 learners' word associations

	Mean USF				**Mean Bespoke**				**Demographics**			
Subject	T1	T2	T3	T4	T1	T2	T3	T4	**gender**	**age**	**L1**	**L1 distance**
2	0.65	0.67	0.31	NA	0.39	0.44	0.19	NA	female	47	Spanish	2.25
7	0.58	0.40	0.43	0.56	0.68	0.32	0.57	0.36	female	33	Azerbaijani	2
22	0.62	0.26	0.50	0.28	0.54	0.11	0.50	0.11	male	34	Spanish	2.5
31	0.56	0.47	0.47	0.27	0.39	0.32	0.35	0.15	female	20	Japanese	1

During initial data collection (Time 1), canonical WAs produced by these subjects included primarily meaning-based associations (see Appendix E); for example, synonyms (e.g., *attack–fight* [S2, T1], *large—big* (S22, T1); antonyms (e.g., *truth—lie* [S7, T1], *warmth—cold* [S7, T1]); and hierarchical sets (e.g., *science—biology* [S31, T1], *wolf—animal* [S2, T1]). Learners also produced canonical WAs that could be categorized as both conceptual (i.e., meaning-based) and collocational (i.e., position-based). These included pairings such as *reflect—mirror* (S2, T1), *game—play* (S22, T1), *liberty—statue* (S31, Time 1), *music—listen* (S7, T1), *send—letter* (S7, T1), *power—strong* (S22, T1), and *basket—ball* (S2, T1). Note that both cue-response (such as *basket—ball*) and response-cue (such as *statue—liberty*) collocations were represented in these WAs.

Analysis at later stages of data collection revealed that the same learners' responses became more difficult to categorize based on existing taxonomies. Overall, learners' responses at Time 3 and 4 were meaning-based, form-based, a combination of meaning and form (specifically, dual-coding), or entirely idiosyncratic (or "hodgepodge"). Many meaning-based associations were personal, such as *lonely—self* (S2, T3), *religion—Muslim* (S7, T4), *pride—my policy* (S31, T4), *family—dear* (S7, T4), and *character—my self* (S31, Time 4). These associations align with previous findings suggesting that L2 speakers may "develop word associations based on feelings, attitudes or strong memories" (Fitzpatrick, 2006, 140).

Examples of form-based associations produced during Time 3 and Time 4 included those based solely on formal features (e.g., *weak—week* [S2, T3]), as well as two-step form associations, such as *sneak—nut* (S7, T3), where the subject presumably produced a meaning-based response due to formal features in the cue (in this case, *sneak* triggered associations for the orthographically similar *snack*). Examples of dual-coding, whereby the response and cue are related in both form and general meaning, included *loss—lost* (S2, T3), *loss—lose* (S31, T4), and *draw—drew* (S22, T3). These results support previous research in which non-native speakers demonstrated significantly more form-based associations than native-speakers (Fitzpatrick, 2006). Because Fitzpatrick's (2006) non-native speaking participants were more advanced than the L2 learners analyzed in the current study (all had experience with academic English at the graduate or post-graduate level), it could be argued that the learners in the current study produced responses at Times 3 and 4 that were less canonical but more indicative of what has been observed among proficient L2 English speakers. Like Fitzpatrick's participants, the L2 learners in the current study were producing more form-based associations and more associations based on "feeling, attitudes or strong memories" (p. 140) by the end of data collection.

Finally, hodgepodge responses produced by the L2 learners analyzed at Time 3 and Time 4 included WAs such as *enjoy—surprise* (S2, T3), *kill—shit* (S31, T4), *advice—go ahead* (S22, T4), *pond—coin* (S7, T3), and *effort—can* (S7, T4). While associative links appear to exist between some of these words (e.g., *go ahead* is a form of advice, someone might throw a coin into a pond the same way coins are thrown into fountains for good luck, to abandon something is to stop), they remain difficult to code and are thus considered "erratic," a subcategory of "hodgepodge" (see Appendix E). However, past research demonstrates that non-native speaker WAs can be somewhat vague (Sökmen, 1993), and Fitzpatrick (2006) has observed a great deal of variation even in native-speaker responses, suggesting that just as each individual's lexicon is unique, each lexical entry may be unique as well, which different saliencies. As with the trend toward more form-based associations, this move toward more idiosyncratic, non-canonical responses observed among L2 learners may suggest that as they develop lexically, learners are behaving more like proficient L2 speakers (i.e., their WAs are exhibiting a great deal of variation) but not necessarily exactly like native-speakers. Alternatively, we may see a boredom effect here, whereby subjects grew tired of the WA task over multiple data collections and thus began to respond more creatively at Times 3 and 4. That said, the production of more creative WAs may also indicate second language play, which in itself might index development (Bell, Skalicky, & Salsbury, 2014; Belz, 2002).

In addition to the main effect of time, learners' self-reported familiarity with the cue word significantly predicted both USF and bespoke canonicity (Tables 4.5 and 4.7), with more familiar cue words resulting in more canonical WAs. This finding is one that is predicted by the literature (e.g., Schmitt & Meara, 1997) and aligns with intuition: Learners would not be expected to produce the same associations as native speakers in response to cue words they don't know well or at all. However, the fact that familiarity was retained as a significant effect in both models—despite prior removal of any WA for which subjects reported little-to-no familiarity with the cue word—suggests that familiarity remains a key indicator of WA performance even for relatively familiar words.

Finally, while the four different sublists assigned randomly to L2 learners across longitudinal data collection did not impact USF canonicity, they did significantly predict bespoke canonicity. This main effect does not change the interpretation of the findings above, but it does suggest that even when stimuli are carefully balanced across stimuli subsets for longitudinal data collection, the subsets should be included in statistical analysis.

One limitation of Study 1 was the relatively small sample size of L2 language learners who contributed meaningful data for analysis. Another limitation was the age of the USF NES word association norms, which were collected in 1973. A more recently collected large-scale dataset of NES word association norms would be ideal and could obviate the need for a locallycollected bespoke dataset in future studies of this nature. Additionally, while L1 distance was included as a control variable, this study failed to account for the influence of learners' L1 on their word associations, which could certainly account for variation in word association behavior (Ecke, 2015; Jarvis & Pavlenko, 2007). This represents a major limitation of the current study and an area for future work investigating L2 word associations and lexical knowledge.

Because results tentatively suggest that the L2 learners in the current study were beginning to produce WAs more like proficient non-native speakers than like NESs, a future study could collect cross-sectional WAs in response to these same stimuli but from a larger sample of L2 learners and functionally bilingual (FB) participants. As is, the number of FB participants in the current study was not large enough for meaningful comparison.

Future investigation of these (or similar) data might also include qualitative analysis of all subjects' WAs, including native and functionally-bilingual subjects, to determine how category-of-response profiles (based on the taxonomy utilized in Appendix E) differ and/or develop longitudinally. Future work could also investigate the lexical and psycholinguistic properties of the WAs produced, such as frequency, concreteness, hypernymy, etc.

4.3 Conclusion

This chapter investigated the development of L2 lexical development in adult learners' English L2 word association (WA) behavior. Analysis centered around the canonicity (or native-likeness) of learners' responses as compared to a large dataset of NES norms and a bespoke dataset collected in parallel to the L2 dataset. Results found that learners produced less canonical WAs over time, though that trend was only significant ($p < .05$) for canonicity derived from the NES norms (Research Questions 1a and 3a). Post-hoc qualitative analysis suggested that the non-canonical WAs produced later during data collection were more personal and idiosyncratic.

Learners also produced more responses that were formally related to the cue word. These findings demonstrate that while L2 WAs do reveal change over time, they don't necessarily become more like native speakers. When contextualized by previous research, results tentatively suggest that the learners analyzed here may have begun producing WAs more like proficient non-native speakers. It may also be that these particular learners are not yet converging on NES norms because they are at earlier stages of development (see Crossley et al., in press, for example of similar findings with frequency).

Neither canonicity measure was impacted by vocabulary size (Research Question 2a): While further inquiry is needed, it may be that vocabulary size has no immediate impact on the sub-construct of lexical network knowledge operationalized by word associations. Finally, it was demonstrated that familiarity with the cue word, even for cues with which learners have indicated a moderate threshold of familiarity, can impact the canonicity of WAs.

5 LEXICAL DECISION SEMANTIC PRIMING: STUDY 2

Lexical decision semantic priming involves participants completing a lexical decision (LD) task in response to a word or pseudoword (spoken or written) that has been preceded by another word. Evidence of priming manifests as faster response times on the LD task when a real word has been preceded by a semantically related word than by an unrelated one (Neely, 1977), with faster latencies in related conditions interpreted as facilitation in processing due to stronger links in the lexicosemantic networks between the two words. The accumulation of evidence from within-language L2 lexical decision semantic priming studies (see Chapter 2) suggests that L2 speakers begin to organize and access semantic information in the same way that L1 speakers do and that the facilitation effects of semantic priming are much stronger for proficient L2 subjects than less proficient ones (Favreau & Segalowitz, 1983; Frenck-Mestre & Prince, 1997; Vasos, 1983). However, to date, attempts to investigate within-language L2 semantic priming longitudinally have not occurred. Nor do researchers know which type of relationship between prime and target is acquired first by L2 learners or whether the relative strength of association type in learners' L2 lexicon changes over time.

The goal of the study is to observe the effects of English semantic priming in L2 learning over time to track the development of receptive L2 associational knowledge. Another goal of the study is to observe the impact of different types of word relationships on both L2 and native/proficient speakers. Stimuli were designed so that the influence of the type of lexicosemantic relationship between primes and targets (e.g., the prime is a synonym of the target or the prime is a collocate of the target) on observed priming effects could be traced over time. This chapter is motivated by the following questions:

1b. Do L2 learners demonstrate evidence of increased L2 semantic priming over time?

2b. What impact does prime-target relationship have on L2 semantic priming and does this influence change over time?

3b. What relation is there between L2 learners' longitudinal priming behavior and vocabulary size?

5.1 Method

The goal of Study 2 (S2) was to observe longitudinal semantic priming effects (SPE) in L2 learners' lexical decisions so as to track the development of L2 lexical knowledge over time. Undergraduate (UG) English speakers' lexical decisions were also observed (cross-sectionally) as a baseline for comparison. S2 was conducted online via the online experiment platform Testable (www.testable.org). The experiment employed a single-trial within-modal unmasked priming task (Neely, 1976). In each trial, the prime word (e.g., *doctor*) appeared on screen for 250 ms and was then automatically replaced by either the target word (e.g., *nurse*) or a pseudoword (e.g., *blerk*). Subjects were instructed to respond to the target word only and pressed "Q" if the character string was indeed a real word in English or "P" if it was a nonsense word in English. In this regard, S2 also incorporated a lexical decision (LD) task.

S2 purposefully manipulated the relationship between prime and target words (e.g., the prime is a synonym of the target, the prime rhymes with the target) so as to determine which associative relationships were most facilitative of language processing and how facilitation changed over time. Prime-target pairs were selected so as to represent an equal amount of meaning-based (e.g., *exit-enter*), position-based (e.g., *department-store*), and form-based associations (e.g., *take-thanks*). These association categories were derived from the word association (WA) cue-response employed in Study 1 (Appendix E). For targets that were not real words in English, pseudowords were used (Rastle et al., 2002). An additional condition included fillers, in which both prime and target were real words but were not associated with one another lexically or semantically. L2 learners took the experiment (with counterbalanced stimuli) once every two months for six months (up to four collections total). UGs subjects took the experiment (all trials) one time.

5.1.1 Participants

S2 analyzed the semantic priming data of L2 learners (n = 21) longitudinally and of 41 native English speakers (NES) and 19 functionally-bilingual (FB) non-native English speakers cross-sectionally. All cross-sectional subjects were undergraduates (UGs). Chapter 3 offers details regarding the recruitment and background of S2 participants.

5.1.2 Stimuli

S2 stimuli consisted of prime and target word pairs that were either both real words in English (e.g., prime = *apple*, target = *letter*) or a real word in English immediately followed by a pseudoword (e.g., prime = *apple*, target= *bletter*). The term *pseudoword* refers to a string of letters that is a not a real word in English (i.e., it is a nonsense word) but still conforms to English phonotactics, resulting in a string of letters that could potentially still be pronounced by a literate English speaker (e.g., *bletter* functions as a pseudoword in English, but *tbtlree* is not).

Four subsets of such stimuli (120 prime-target pairs each; 480 total; see Table 5.1, Appendix F, and description of conditions below) were randomly assigned to L2 learners across four different data collection periods. Meanwhile, UG (i.e., cross-sectional) subjects each took exactly half of the total Study 2 stimuli: half of UG subjects saw the first 240 prime-target pairs (i.e., subset L2_A and L2_B) while a second half of UG subjects saw the other 240 (i.e., subset L2_C and L2_D).

Table 5.1 Study 1 Semantic Priming Stimuli

		Real word targets				Pseudoword target	Total Pairs	
Subsets		Meaning-based prime	Position-based prime	Form-based prime	Unrelated prime	Real word prime	L2	UG
L2_A	UG_1	10	10	10	30	60	120	240
L2_B		10	10	10	30	60	120	
L2_C	UG_2	10	10	10	30	60	120	240
L2_D		10	10	10	30	60	120	
Totals		40	40	40	120	240	480	

Note. All primes in Study 2 were real words.

5.1.2.1 Word Pair Conditions

The real-word condition in S2 was further broken into four sub-conditions based on the associative relationship between the prime and the target: a meaning-based relationship, a position-based relationship, a form-based relationship, or no associative relationship at all (see Table 5.1). The first three of these subconditions based on association relationships were further broken into micro-relations (see Table 5.2) derived from the taxonomy of word associations utilized in Study 1 (see Appendix E). These micro-relations were not analyzed statistically as dependent variables in S2; rather, they were balanced across studies to ensure that the more general associative relations analyzed (e.g., meaning-based) contained the full range of association types that might constitute that more general category. Meanwhile, unrelated real-word pairs (i.e., no associative relationship at all, e.g., *security* and *diet*) were intended to serve as a control and to ensure that subjects did not guess the purpose of the experiment. Unrelated pairs constituted 50% of stimuli, with related pairs (50% total) divided evenly among meaning-based (16.7%), position-based (16.7%), and form-based (16.7%) pairs.

*Table 5.2 Study 2 Real Word Relations per Subset Stimuli**

Relation	Micro-relation	Number	Example
Meaning-based	Synonym	2	*dumb, stupid*
	Antonym	2	*late, early*
	Conceptual Other	2	*bee, honey*
	Coordination	2	*square, circle*
	Hierarchical Set	2**	*color, red*
	Total	*10*	
Position-based	Cue-response	7	*brand, name*
	Response-Cue	3	*juice, orange*
	Total	*10*	
Form-based	Affix	2	*loosen, loose*
	Formal	6	*deal, meal*
	Two-step	2	*stake, sauce*
	Total	*10*	
Unrelated	n/a	30	*glass, thumb*
	Total	*30*	

** Stimuli broken into four sets for L2 subjects; UG subjects encountered two of such sets at a time*
*** One pair was hyponym-hypernym, and the other was hypernym-hyponym*

All real-word prime-target word pairs used in S2 stimuli (with the exception of the unrelated condition) were selected from the University of South Florida (USF) NES word association norms (http://w3.usf.edu/FreeAssociation) using cue-to-target forward strength (FSG). FSG values were calculated by USF researchers as the proportion of subjects who produced a given target in the presence of a cue word. Thus, FSG scores describe the relationship between a cue and a target, rather than describing the associative properties of either word in isolation. For example, the FSG between *dagger* and *knife* is .614, meaning that when presented with the cue word *dagger* in a word association task, 61% of NESs produced the word *knife*.

However, the FSG between *dagger* and *murder* is .010, meaning that only 10% of NES subjects produced the word *murder* in response to the target *dagger*.

For meaning-based and position-based stimuli, all prime-target pairs had an FSG $\geq .20$ (meaning that over 1/5 of respondents produced the target when presented with the cue/prime) and $\leq .80$ (thus avoiding situations where over 4/5 of respondents produced the target). These selection criteria ensured that the target had a strong associative relationship in NESs' minds to the prime word while avoiding extremely strong associations that might be overly-deterministic as stimuli.

Because form-based responses (e.g., *deal, meal*) in NES word association data are rarer, form-based prime-target associations were selected from a subset of more idiosyncratic USF responses. Due to their idiosyncratic nature, these pairs had much lower FSGs (< 0.01). Still, all were produced by at least one NES in a word association task. FSG values for all word pairs were also included as fixed factors in analyses.

Word pairs for the unrelated condition were created by recycling primes and targets from other subsets one time each. This was done in such a way to ensure that no subject saw the same word twice in the same experiment. Word pairs in the unrelated condition were reviewed independently by three researchers to ensure that none of the primes and targets bore a meaning-, position-, or form-based relationship to one another.

5.1.2.2 Real words

All prime and target real words were selected from the top 5,000 lemmas in Corpus of Contemporary American English (COCA, http://www.wordfrequency.info/free.asp); thus, every real word used in S2 (regardless of its status as a target or a prime) was among the first 5,000 most frequent words in American English. This was done to increase the likelihood that L2 learners subjects were already familiar with the real words they encountered during the experiment.

Real-word targets were controlled statistically across conditions for both frequency and length. Descriptive statistics for real-word target length and frequency are reported in Tables 5.3 and 5.4. No significant differences were found among the four conditions (meaning, position, form, and unrelated) for word length, $F(3, 236) = 0.538$, $p = .664$, or frequency, $F(3, 236) = 0.042$, $p = .998$. Nor were there significant differences in length or frequency when conditions were compared across substudies (e.g., there were no significant differences between the length of words in the meaning-based condition in substudy 1 when compared to the length of words in the meaning-based condition in substudy 2; see Tables 5.5 and 5.6, respectively).

Targets were also balanced for part of speech across sub-sets. Within the meaning-based and form-based conditions, real-word targets in each sub-study contained two adjectives, two verbs, and six nouns per condition. Owing to the nature of position-based cues and targets in word association data, all real-word targets in the position-based condition were nouns. Meanwhile, real-word primes in the pseudoword condition contained 40 adjectives, 40 verbs, and 160 nouns total, with 10 adjectives, 10 verbs, and 40 nouns per sub-study (see Appendix F). Primes for the pseudoword condition were selected to roughly match the frequency and length of other real words in S2. Like all other real words, primes from the pseudoword condition were selected from the first 5,000 lemmas in COCA, and none are repeated in other conditions within S2.

Table 5.3 Study 2 Length of Targets by Condition, Descriptive Statistics

Prime-Target Relation	**Substudy**	**N**	**Mean length**	**Std. Dev**	**95% Confidence Interval for Mean**		**Min**	**Max**
Meaning-based	1	10	4.700	1.252	3.805	5.595	3	7
	2	10	4.800	1.317	3.858	5.742	3	8
	3	10	4.400	0.966	3.709	5.091	3	6
	4	10	4.700	1.160	3.871	5.529	3	7
	Total	40	4.650	1.145	4.284	5.016	3	8
Position-based	1	10	5.000	0.943	4.326	5.674	4	7
	2	10	4.500	0.850	3.892	5.108	3	6
	3	10	5.000	2.000	3.569	6.431	3	8
	4	10	4.900	0.994	4.189	5.611	4	7
	Total	40	4.850	1.252	4.450	5.250	3	8
Form-based	1	10	4.500	0.707	3.994	5.006	4	6
	2	10	5.000	1.563	3.882	6.118	3	8
	3	10	5.000	0.816	4.416	5.584	4	6
	4	10	4.900	1.197	4.044	5.756	3	6
	Total	40	4.850	1.099	4.499	5.201	3	8
Unrelated	1	30	4.733	1.081	4.330	5.137	3	7
	2	30	4.767	0.898	4.431	5.102	3	6
	3	30	4.767	1.251	4.300	5.234	3	8
	4	30	4.867	1.042	4.478	5.256	3	7
	Total	**120**	4.783	1.063	4.591	4.975	3	8

Table 5.4 Study 2 Frequency of Targets by Condition, Descriptive Statistics

Prime-Target Relation	Subset	N	Mean freq	Std. Dev	95% Confidence Interval for Mean		Min	Max
Meaning-based	1	10	48990.800	33663.117	24909.657	73071.943	7742	108171
	2	10	46030.400	38362.604	18587.446	73473.354	13541	133571
	3	10	47016.900	26551.140	28023.358	66010.442	8117	86184
	4	10	48957.000	43974.312	17499.672	80414.328	9153	150718
	Total	40	47748.775	34809.847	36616.046	58881.504	7742	150718
Position-based	1	10	47302.000	41487.574	17623.578	76980.422	9755	127139
	2	10	44875.500	47863.789	10635.808	79115.192	11446	135986
	3	10	45498.300	30260.628	23851.151	67145.449	7694	86231
	4	10	43997.200	27761.787	24137.614	63856.786	11846	87427
	Total	40	45418.250	36284.467	33813.914	57022.586	7694	135986
Form-based	1	10	46311.700	51138.972	9729.083	82894.317	10324	152891
	2	10	44981.600	32809.651	21510.990	68452.210	15260	123183
	3	10	42871.400	28461.962	22510.939	63231.861	13769	114094
	4	10	47584.600	47912.203	13310.274	81858.926	9212	150646
	Total	40	45437.325	39645.105	32758.205	58116.445	9212	152891
Unrelated	1	30	42783.033	30349.797	31450.233	54115.834	5855	114094
	2	30	45213.700	41488.151	29721.770	60705.630	5434	155032
	3	30	44959.100	39012.987	30391.411	59526.789	11446	135986
	4	30	46409.433	41527.772	30902.709	61916.158	7742	152891
	Total	120	44841.317	37905.773	37989.562	51693.072	5434	155032

Table 5.5 Study 2 Comparing Length of Targets in Conditions by Substudy

	Test of Homogeneity of Variances				ANOVA						Welch's Robust Test of Equality of Means			
Prime-Target Relation	**Levene**	**df 1**	**df 2**	**Sig.**		**Sum of Sqrs**	**df**	**Mean Square**	**F**	**Sig.**	**Stat**	**df 1**	**df 2**	**Sig.**
Meaning-based	0.091	3	36	0.965	Between groups	0.9	3	0.3	0.215	0.885	0.248	3	19.860	0.862
					Within groups	50.2	36	1.394						
					Total	51.1	39							
Position-based	3.891	3	36	0.017*	Between groups	1.7	3	0.567	0.343	0.794	0.591	3	19.86	0.628*
					Within groups	59.4	36	1.65						
					Total	61.1	39							
Form-based	3.154	3	36	0.037*	Between groups	1.7	3	0.567	0.449	0.719	0.806	3	19.352	0.506*
					Within groups	45.4	36	1.261						
					Total	47.1	39							
Unrelated	0.477	3	116	0.699	Between groups	0.3	3	0.1	0.087	0.967	0.089	3	64.049	0.966
					Within groups	134.067	116	1.156						
					Total	134.367	119							

Table 5.6 Study 2 Comparing Frequency of Targets in Conditions by Subset

	Test of Homogeneity of Variances				ANOVA						Welch's Robust Test of Equality of Means			
Prime-Target Relation	**Levene Statistic**	**df 1**	**df 2**	**Sig.**		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Stat**	**df 1**	**df 2**	**Sig.**
Meaning -based	0.576	3	36	0.634	Between groups	6.49E+07	3	2.16E+07	0.017	0.997	0.015	3	19.67	0.997

					Within groups	4.72E+10	36	1.31E+09						
					Total	4.73E+10	39							
Position-based	1.682	3	36	0.188	Between groups	5.87E+07	3	1.96E+07	0.014	0.998	0.014	3	19.623	0.998
					Within groups	5.13E+10	36	1.42E+09						
					Total	5.13E+10	39							
Form-based	2.369	3	36	0.087	Between groups	1.22E+08	3	4.06E+07	0.024	0.995	0.028	3	19.512	0.994
					Within groups	6.12E+10	36	1.70E+09						
					Total	6.13E+10	39							
Un-related	0.855	3	116	0.467	Between groups	2.05E+08	3	6.85E+07	0.047	0.987	0.056	3	63.85	0.982
					Within groups	1.71E+11	116	1.47E+09						
					Total	1.71E+11	119							

In addition to controlling statistically for word length and frequency in stimuli design, length and frequency, as well as other lexical and psycholinguistic properties of the targets (specifically word length, familiarity, concreteness, and number of associates in L1 WA norms), were included as fixed effects during post-hoc analyses (see below).

5.1.2.3 Pseudowords

Pseudowords used in S2 were extracted from the English Lexicon Project (Balota et al., 2007). Pseudowords were selected to match the word length distribution of real-word primes for the pseudo-word condition exactly. The final pseudoword list was checked by three independent researchers to ensure that no pseudoword closely approximated an actual word (including proper nouns or slang) in English. For the complete set of S2 stimuli, including real-word primes, and real- and pseudoword targets, see Appendix F.

5.1.3 Procedure

Data collection was facilitated using the online experiment platform Testable (www.testable.org). The experiment employed a single-trial within-modal unmasked priming task (Neely, 1976). Following Balota et al. (2007), a fixation point was presented for 250 ms prior to each prime word. The prime appeared on screen for 250 ms and was then replaced by the target following a 200 ms inter-stimulus interval. Subjects were instructed to respond to the target (i.e., the second) word only and pressed “Q” if the character string was indeed a real word in English or “P” if it was a nonsense word in English. Participant reaction time (measured in milliseconds) and accuracy was recorded for each trial. Each L2 learner completed 120 trials (60 real words, 60 pseudowords) per data collection (up to 480 trials over the course of the study) and each UG participant completed 240 trials (120 real words, 120 pseudowords) in one single data collection. Prior to beginning the experiment, each subject received instructions illustrating the task and 8 practice trials.

5.1.4 Analysis

Results from S2 were analyzed in order to determine the following: 1) the relative strengths of prime-target association types (i.e., conditions) in facilitating priming (both L2 learner and UG data analyzed); 2) whether evidence of semantic priming effects (SPE) increased over time (L2 learner data only); and 3) what relationship the acquisition of prime-target relationships (as evidenced by SPE) had to time (L2 learner data only) and vocabulary size (both L2 learner and UG data). Linear mixed effects (LME) models were used to investigate the role of association type and time (i.e., Data collection 1, 2, 3, or 4) in predicting response latencies and accuracies on the LD task while controlling for other effects (e.g., trial number, vocabulary size test scores, demographic info, stimuli sublist) as well.

Following practice in the field (e.g., Balota et al., 2007), a two-step outlier identification and removal process was followed. In the first step, observations with RTs ≤ 200ms and ≥ than 3000ms were excluded. In the second step, RTs were standardized within subjects and any observation for which the response time was less than three standard deviations (SDs) below the mean for a given subject or greater than three SDs above the mean for that subject was removed.

After removal of outlier trials, linear mixed effects models were used to predict response latencies on the lexical decision task. Separate models for the longitudinal L2 learner and cross-sectional UG were constructed. For the L2 learner model, gender, handedness, age, L1 distance[11], VLT score, self-reported proficiency, exposure (mean score based on survey results), average hours of English used per day, trial order, stimuli sublist, stimuli FSG, target

---

[11] L1 distance ranges from 1 to 3 (with a score of 3 reflecting a language very close to English) and is calculated based on the difficulty of learning L2 English as a function of one's L1 (Chiswick & Miller, 2005).

concreteness, target frequency, target length, time, and relatedness condition (whether the prime and target were unrelated or had a form-based, meaning-based, or position-based relationship to one another) were included as fixed effects. For the categorical variables, the following baselines were set: *gender*: female; *handedness*: left; *time*: Time1; *relatedness*: unrelated. Subjects and items were entered as random effects, with a random slope for relatedness condition added to subjects. Interactions between time and the relatedness condition and between VLT and relatedness were investigated as well.

For the NES and FB models, the same fixed effects and categorical baselines as the L2 learner model were entered[12], with the exception of time and variables related to L2 learning (specifically, L1 distance, self-proficiency, exposure, and average hours of English used per day). Because the inclusion of VLT in UG analysis caused the LME model to fail to converge and to drop participant data—and because initial models demonstrated that vocabulary size was not a significant main effect in either the NES or FB models—VLT was removed from UG models in order to maintain a larger sample size.

5.2 Results and discussion

5.2.1 Descriptive statistics

Prior to outlier removal, there were 4,244 accurate observations in response to real words in the L2 learner data and 7,109 in the UG data. The first outlier removal step removed 355 observations in the L2 learner data and 91 observations in the UG data. In the second stage of outlier removal, any observation for which the RT was ≤ -3 or ≥ 3 SDs of the mean RT for a given subject was removed. This second step removed 80 observations in the L2 data and 149 observations in the UG data. In the L2 learner dataset, 3808 observations remained for analysis,

[12] In the NES data, an additional gender category (other) was reported, though the baseline was still set to female.

while 4618 NES and 2134 FB observations remained for UG analysis. After this trimming, the RTs remained positively skewed across both datasets, so a transformation (the base-10 logarithm of each RT) was performed on the RTs for all subjects' UG data.

Table 5.7 displays descriptive statistics for all participant response times using raw (i.e., non-transformed) RTs. The table is organized according to the relationship between the prime and the target: unrelated vs. related, with the related condition further broken into meaning- , position -, or form-based relationships[13].

5.2.2 L2 Learner LME

An initial LME model predicting RTs for L2 learners found no significant interactions between VLT and relatedness, so a second model was refit retaining only the significant interaction between time and relatedness. The resulting LME model is reported in Table 5.8, which displays the coefficients, standard error, *t* values, and *p* values for the L2 learner model.

Significant ($p < 0.05$) and approaching significant ($p < 0.10$) variables from the model in Table 5.8 were then entered into the final LME model reported in Table 5.9. This final model predicting RTs for L2 learners found no significant difference in RTs between the unrelated and related prime-target condition (form-based, meaning-based, or position-based) overall, nor was there a significant interaction between time and this condition (see Table 5.9).

Exposure, self-reported proficiency, average number of hours spent speaking English per day (i.e., *use*), frequency of the target word, trial order, sublist, and time were significant predictors. This final L2 learner model reported a marginal R^2 of 0.052 and a conditional R^2 of 0.406. Visual inspection suggested the model was not impacted by heteroscedasticity. Table 5.9 displays the coefficients, standard error, *t* values, and *p* values for the final L2 learner model.

[13] Initial models found no significant differences when related pairs as a whole were directly compared to unrelated pairs and not further analyzed as meaning- , position-, or form-based relationships.

Table 5.7 Study 2 Mean and Standard Deviations for Response Times for All Participants

	L2			Functionally Bilingual (Undergrad)			NES (Undergrad)		
	M	SD	N	M	SD	N	M	SD	N
Related	770.222	329.119	1920	628.086	216.411	1065	719.683	256.449	2314
Meaning-based	748.614	303.251	652	635.387	235.058	344	703.533	251.629	762
Position-based	774.458	336.121	649	620.786	217.212	359	713.395	260.057	780
Form-based	788.543	346.656	619	628.390	196.585	362	741.977	256.274	772
Unrelated	787.742	340.872	1888	644.746	230.535	1069	720.941	262.838	2304

Note. Table displays raw, non-transformed reaction time data with outliers (<200 and <3000ms; <3 and >3 per-subject SDs) and incorrect responses removed.

Table 5.8 Study 2 Linear Mixed Effects Model Predicting Reaction Time for L2 Learners, All Variables

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**2.930**	**0.202**	**14.495**	**< 0.001**
Demographic Variables				
Age	0.001	0.003	0.249	0.806
Gender: male	-0.015	0.061	-0.255	0.802
Hand: right	-0.034	0.131	-0.264	0.795
Individual Differences				
Hours per day*	**-0.010**	**0.003**	**-3.442**	**0.001**
L1 distance	-0.009	0.060	-0.149	0.884
Self-reported proficiency*	**0.037**	**0.006**	**6.441**	**< 0.001**
Exposure*	**-0.003**	**0.001**	**-2.600**	**0.009**
VLT	0.000	0.000	-1.131	0.258
Linguistic Features				
Concreteness (target)	-0.003	0.003	-0.739	0.461
Frequency (target)*	**-0.020**	**0.006**	**-3.572**	**< 0.001**
Length (target)	0.002	0.003	0.673	0.502
Stimuli FSG (pair)	0.000	0.000	-0.530	0.596
Procedural Variables				
Trial order*	**0.000**	**0.000**	**-6.696**	**< 0.001**
Sublist (baseline: Sublist 1)				
Sublist 2*	**0.021**	**0.007**	**3.138**	**0.002**
Sublist 3*	**-0.023**	**0.012**	**-2.006**	**0.046**
Sublist 4	0.016	0.012	1.347	0.179
Experimental Variables				
Semantic relationship (baseline: Unrelated)				
Related: form	0.014	0.015	0.969	0.333
Related: meaning	-0.007	0.017	-0.408	0.684
Related: position	-0.004	0.017	-0.264	0.792
Time (baseline: Time 1)				
Time2*	**0.052**	**0.010**	**5.360**	**< 0.001**
Time3	0.020	0.010	1.955	0.051
Time4	-0.005	0.011	-0.468	0.640
Condition: Time x Semantic Relationship				
Time2 x form	-0.015	0.019	-0.762	0.446

Time3 x form	-0.038	0.020	-1.906	0.057
Time4 x form	-0.010	0.021	-0.473	0.636
Time2 x meaning	0.002	0.018	0.105	0.916
Time3 x meaning	-0.005	0.019	-0.243	0.808
Time4 x meaning	0.014	0.021	0.660	0.510
Time2 x position	0.026	0.019	1.388	0.165
Time3 x position	0.000	0.019	-0.024	0.981
Time4 x position	0.011	0.021	0.510	0.610

Note: Reaction times were transformed using base-10 logarithm of each RT
**Significant predictor of reaction times.*

Table 5.9 Study 2 Final LME Model Predicting Reaction Time for L2 Learners, Significant Variables Only

Fixed Effect				
(Intercept)*	**2.896**	**0.035**	**82.617**	**< 0.001**
Individual Differences				
Hours per day*	**-0.010**	**0.003**	**-3.441**	**0.001**
Self-reported proficiency*	**0.035**	**0.006**	**6.292**	**< 0.001**
Exposure*	**-0.002**	**0.001**	**-2.429**	**0.015**
Linguistic Features				
Frequency (target)*	**-0.021**	**0.005**	**-4.026**	**< 0.001**
Procedural Variables				
Trial order*	**0.000**	**0.000**	**-6.721**	**< 0.001**
Sublist (baseline: Sublist 1)				
Sublist 2*	**0.019**	**0.006**	**3.032**	**0.003**
Sublist 3*	**-0.024**	**0.011**	**-2.091**	**0.038**
Sublist 4	0.016	0.011	1.412	0.159
Experimental Variables				
Semantic relationship (baseline: Unrelated)				
Related: form	0.015	0.015	1.063	0.288
Related: meaning	-0.012	0.014	-0.861	0.390
Related: position	-0.010	0.014	-0.733	0.464
Time (baseline: Time 1)				
Time2*	**0.050**	**0.010**	**5.243**	**< 0.001**
Time3	0.018	0.010	1.810	0.070
Time4	-0.007	0.011	-0.653	0.514

Condition: Time x Semantic Relationship				
Time2 x form	-0.015	0.019	-0.793	0.428
Time3 x form	-0.038	0.020	-1.931	0.054
Time4 x form	-0.010	0.021	-0.474	0.636
Time2 x meaning	0.002	0.018	0.107	0.915
Time3 x meaning	-0.005	0.019	-0.237	0.813
Time4 x meaning	0.014	0.021	0.656	0.512
Time2 x position	0.026	0.019	1.384	0.167
Time3 x position	-0.001	0.019	-0.028	0.978
Time4 x position	0.010	0.021	0.488	0.626

Note: Reaction times were transformed using base-10 logarithm of each RT
**Significant predictor of reaction times.*

Posthoc comparisons found no significant interactions between time and semantic relatedness when time was re-leveled; however, there were significant interactions for Time alone on RTs between Time 1 and 2; Times 2 and 3; Times 2 and 4; and Times 3 and 4 (Table 5.10).

On the whole, learners with higher self-reported proficiency resulted in slower reaction times, while learners who reported greater exposure to and use of English obtained quicker reaction times (regardless of the relationship, if any, between the prime and target). Higher word frequency also resulted in quicker reaction times, as did stimuli that occurred earlier in the experiment (i.e., trial order). A significant effect was also found for the stimuli sublist in which a word occurred. Finally, real words in English were recognized more slowly at Time 2 than at any other time, though this effect for time did not interact with the semantic relationship (or lack thereof) between the prime and target.

Table 5.10 Post-hoc Comparisons of Time on Reaction Times

Fixed Effect	Coefficient	Std. error	*t*	*p*
Time 1 vs. TIME 2*	**0.050**	**0.010**	**5.243**	**< .001**
Time 1 vs. Time 3	0.018	0.010	1.810	0.070
Time 1 vs. Time 4	-0.007	0.011	-0.653	0.514
TIME 2 vs. Time 3*	**-0.032**	**0.010**	**-3.194**	**0.001**
TIME 2 vs. Time 4*	**-0.058**	**0.011**	**-5.116**	**< .001**
TIME 3 vs. Time 4*	**-0.026**	**0.011**	**-2.335**	**0.020**

	Time 1		**Time 2**		**Time 3**		**Time 4**	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
RT z-scores lg10	2.832	0.188	2.881	0.176	2.856	0.155	2.847	0.157
RT raw	744.355	329.014	829.073	381.939	767.676	307.891	749.921	287.559

5.2.3 NES Data LME

An initial LME model predicting RTs for NES UG subjects was conducted as a baseline for this data collection method and stimuli. The NES model found significant main effects for frequency and trial order[14]. Table 5.11 displays the coefficients, standard error, *t* values, and *p* values for this initial UG model.

Table 5.11 Study 2 Linear Mixed Effects Model Predicting Reaction Times for NES Only

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**2.845**	**0.051**	**55.825**	**< .001**
Demographic Variables				
Age	0.001	0.001	1.112	0.274
Gender: male	0.038	0.032	1.209	0.235
other	0.004	0.070	0.054	0.957
Hand (baseline:right)	-0.029	0.044	-0.667	0.509
Linguistic Features				
Concreteness (target)	-0.004	0.002	-1.954	0.052

[14] An initial UG model found no significant differences when related pairs as a whole were directly compared to unrelated pairs and not further analyzed as meaning- , position-, or form-based relationships.

Frequency (target)*	**-0.013**	**0.003**	**-3.875**	**0.000**
Length (target)	0.002	0.002	0.999	0.319
Stimuli FSG (pair)	0.000	0.000	0.217	0.828
Procedural Variables				
Trial order*	**0.000**	**0.000**	**-3.076**	**0.002**
Sublist (baseline: Sublist 1)	-0.019	0.031	-0.597	0.555
Experimental Variables				
Condition: Semantic relationship				
Related: form	0.009	0.005	1.685	0.093
Related: meaning	-0.010	0.007	-1.352	0.178
Related: position	-0.004	0.007	-0.538	0.591

Note: Reaction times were transformed using base-10 logarithm of each RT
**Significant predictor of reaction times.*

Significant ($p < 0.05$) and approaching significant ($p < 0.10$) main effects from the model in Table 5.11 were then re-entered into LME models until only significant variables remained. The resulting model is reported in Table 5.12. This final NES model found a significant ($p < 0.05$) difference between cue-target pairs related via formal association and the unrelated condition, as well as a trend approaching significance ($p < 0.10$) between meaning-based associations and the unrelated condition. NES subjects' reaction times were faster for targets that bore a meaning-based relationship to the cue (e.g., *late-early)*, while word recognition was obstructed (i.e., slowed down) by formally associated targets (e.g., *unite-unison*). Frequency and trial order were main effects as well, with higher target word frequency resulting in quicker reaction times, as well as stimuli that occurred later in the experiment (i.e., trial order). This final model reported a marginal R^2 of 0.007 and a condition R^2 of 0.420. Visual inspection suggested the model was not impacted by heteroscedasticity. Table 5.12 displays the coefficients, standard error, *t* values, and *p* values for the final NES model.

Table 5.12 Study 2 Linear Mixed Effects Model Predicting Reaction Time for NES Only, Significant Variables Only

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**2.889**	**0.018**	**163.148**	**< .001**
Linguistic Features				
Frequency (target)*	**-0.013**	**0.003**	**-4.247**	**0.000**
Procedural Variables				
Trial order*	**0.000**	**0.000**	**-3.077**	**0.002**
Experimental Variables				
Condition: Semantic relationship				
Related: form*	**0.011**	**0.005**	**2.213**	**0.028**
Related: meaning	-0.009	0.005	-1.819	0.071
Related: position	-0.005	0.005	-0.978	0.330

Note: Reaction times were transformed using base-10 logarithm of each RT
**Significant predictor of reaction times.*

5.2.4 FB Data LME

An initial LME model predicting RTs for FB UG subjects was also conducted to compare to L2 learners' lexical network knowledge. An initial FB model found a significant main effect for trial order[15]. Table 5.13 displays the coefficients, standard error, *t* values, and *p* values for this initial FB model.

Table 5.13 Study 2 Linear Mixed Effects Model Predicting Reaction Times for FB UGs Only

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**2.884**	**0.189**	**15.278**	**0.000**
Demographic Variables				
Age	-0.001	0.008	-0.139	0.891
Gender: male	-0.004	0.036	-0.117	0.909
Linguistic Features				
Concreteness (target)	-0.004	0.003	-1.429	0.155

[15] An initial UG model found no significant differences when related pairs as a whole were directly compared to unrelated pairs and not further analyzed as meaning- , position-, or form-based relationships.

Frequency (target)	-0.009	0.005	-1.897	0.060
Length (target)	-0.001	0.003	-0.479	0.633
Stimuli FSG (pair)	0.000	0.000	-1.548	0.124
Procedural Variables				
Trial order*	**0.000**	**0.000**	**-2.222**	**0.026**
Sublist (baseline: Sublist 1)	0.001	0.034	0.037	0.971
Experimental Variables				
Condition: Semantic relationship				
Related: form	-0.013	0.007	-1.713	0.089
Related: meaning	0.005	0.011	0.470	0.640
Related: position	-0.003	0.010	-0.250	0.803

Note: Reaction times were transformed using base-10 logarithm of each RT
**Significant predictor of reaction times.*
Handedness removed bc all participants right-handed

Significant ($p < 0.05$) and approaching significant ($p < 0.10$) main effects from the model in Table 5.13 were then re-entered into LME models until only significant variables remained. The resulting model is reported in Table 5.14. This final FB model found a significant ($p < 0.05$) difference between cue-target pairs related via position-based association and the unrelated condition, whereby subjects' accurate word recognition was facilitated by targets that bore a position-based relationship to the cue (e.g., *mouth-wash)*. Trial order was a main effect as well, with stimuli that occur later in the experiment resulting in quicker reaction times. This final model reported a marginal R^2 of 0.004 and a conditional R^2 of 0.243. Visual inspection suggested the model was not impacted by heteroscedasticity. Table 5.14 displays the coefficients, standard error, *t* values, and *p* values for the final NES model.

Table 5.14 Study 2 Linear Mixed Effects Model Predicting Reaction Times for FB Only, Significant Variables Only

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**2.800**	**0.014**	**193.689**	**< .001**
Procedural Variables				

Trial order*	**0.000**	**0.000**	**-2.222**	**0.026**
Experimental Variables				
Condition: Semantic relationship				
Related: form	-0.011	0.007	-1.544	0.125
Related: meaning	-0.007	0.009	-0.841	0.410
Related: position*	**-0.016**	**0.007**	**-2.289**	**0.024**

Note: Reaction times were transformed using base-10 logarithm of each RT
**Significant predictor of reaction times.*

5.2.5 Discussion

The goal of Study 2 was to observe the effects of L2 semantic priming in learners over time to track the development of receptive L2 associational knowledge. Another goal of the study was to observe the impact of different types of word relationships on L2 learners' semantic priming. A lexical decision task was analyzed under the assumption that priming would manifest as decreased word recognition latencies between a related prime and target when compared to unrelated pairs (Neely, 1977). Results demonstrated no statistical evidence of semantic priming in L2 English learners' lexical decision behavior at any point during longitudinal data collection, regardless of the type of relationship between prime and target (form-based, meaning-based, or position-based). However, some evidence of semantic priming was observed in baseline UG data. Specifically, there was a trend approaching significance among NES subjects whereby meaning-based associations between primes and targets (e.g., *vegetable-fruit*) seemed to facilitate word recognition. Secondly, a significant trend emerged whereby targets that were formally related to cues (orthographically or phonetically; e.g., *opinion-onion*) interfered with word recognition (Table 5.12). For FB subjects, position-based associations between primes and targets (e.g., *phone-bill*) facilitated priming.

Variables that resulted in quicker reactions times in L2 learners' lexical decisions overall (independent of priming) included word frequency of the target, learners' exposure to English, the average number of hours spent speaking English per day, and trial order. Conversely, L2 learners who reported higher self-proficiency resulted in slower reaction times. Despite being balanced for frequency, length, part of speech, cue-to-target strength, etc., stimuli sublist was another significant predictor of L2 RTs. Time was also a significant—albeit nonlinear—predictor of RTs. Table 5.10 indicates that learners were slowest to accurately recognize real words in English at Time 2, a little faster at Time 3, and fastest at Times 1 and 4 (with no statistically significant difference between Time 1 and Time 4). For NES subjects, target word frequency and trial order resulted in quicker reaction times. For FB subjects, only trial order was significant.

Given that past L2 semantic priming studies have demonstrated stronger facilitation effects of semantic priming for more proficient L2 subjects and bilinguals (Frenck-Mestre & Prince, 1997; Kotz, 2001; Kotz & Elston-Güttler, 2004; Schwanenflugel & Rey, 1986; Vasos, 1983), it may be that the L2 English learners analyzed in the current study had not obtained a level of English proficiency necessary for semantic priming to manifest, even after six months of English study in the United States. To the degree that priming effects are "indicative of the strength and richness of semantic relationships among words in a language (McDonough & Trofimovich, 2009, p. 67), one interpretation of these findings is that the L2 learners analyzed had not yet developed such relationships among the words in their L2 lexicon. In order to investigate whether semantic priming might have been just beginning to emerge at the end of L2 data collection, a post-hoc analysis investigated L2 learners' lexical decision behavior at Time 4 only. Results of this post-hoc analysis revealed no significant differences among relatedness conditions, indicating that there were no statistical semantic priming effects even at the end of data collection.

While semantic priming effects were not statistically demonstrated in L2 learners' data, it is worth noting that descriptive statistics of learners' RTs ran parallel to the statistical trend found in NES data (Table 5.7). For both subject groups, formally-associated prime-target pairs seemed to obstruct (rather than facilitate) word recognition, resulting in the slowest RTs overall for this sub-condition—even slower than the unrelated condition. Following the unrelated condition, the next fastest condition was position-based associations, and the fastest (for both L2 and NES) was meaning-based. While these RTs patterns tentatively suggest that L2 learners are beginning to respond to words in a fashion similar to NES, it may be that they have not yet developed the automaticity in L2 word recognition that would result in actual facilitation of lexical processing (Segalowitz & Hulstijn, 2005).

It is not immediately clear why FB data followed a different trend, with position-based association demonstrating priming effects but not meaning-based. Unlike NES, FB subjects' lexical decisions responses also were not affected by form-based prime-target associations, which may have obstructed latencies for L2[16] and NES subjects. It could be that as L2 learners develop proficiency they become more attuned to position-based (i.e., syntagmatic) relationships between lexical items in their L2. However, given that the cross-sectional FB sample size in this study was relatively small (N = 19), further work is needed to corroborate and interpret these findings. Specifically, larger sample size are needed to better represent this population.

A potential explanation for the lack of evidence of semantic priming among L2 learners is stimuli design, as previous research has suggested that L2 semantic priming may not manifest due to the type of relationships analyzed (e.g., Crossley, 2013; Crossley & Skalicky, in press) or weak associations between primes and targets (e.g., Devitto & Burgess, 2004). However, the

[16] Though recall that L2 reaction times between form-based and the unrelated condition were not statistically significant.

current study's stimuli were designed to include a range of association types (hyponyms, synonyms, collocations, category membership, etc.) and were based on cue-to-target forward strength (FSG) derived from attested native-speaker word association norms (see Appendix E and Table 5.2), so it is not immediately clear how the stimuli may have failed to facilitate lexical access.

It could also be that Testable, the self-contained website used to collect online response time data for Study 2, simply fails to record precise enough reaction times to capture semantic priming. However, some evidence of semantic priming was observed among UG data, and other findings from Study 2 align with predictions of (non-primed) lexical decision behavior, suggesting that Testable was indeed sophisticated enough to capture lexical decision latencies regardless of whether word recognition was primed by previous linguistic stimuli. For example, factors contributing to quicker word recognition for both L2 and NES included trial order (i.e., at what point in the experiment a stimulus pair occurred) and the frequency of the target word, with higher frequency words resulting in quicker reaction times[17]. The significance of trial order suggests that subjects began responding to words more quickly as they moved through the experiment, likely due to a practice effect common to the lexical decision task (Forbach, Stanners, & Hochhaus, 1974). Similarly, the finding that higher frequency targets were recognized more quickly (regardless of whether or not the prime was related to said target) aligns with previous research demonstrating that word recognition is impacted by word frequency, both for L1 (Balota et al., 2004; Rubenstein, Garfield, & Millikan, 1970; Scarborough, Cortese, & Scarborough, 1977) and L2 subjects (Berger, Crossley, & Skalicky, in press; Duyck, Vanderelst, Desmet, & Hartsuiker, 2008; Ratcliff, Gomez, & McKoon, 2004). In fact, Duyck et al. (2008)

---

[17] A post-hoc analysis found that frequency was a significant predictor of L2 learners' overall lexical decision performance even at Time 1.

found that bilinguals demonstrated a significantly stronger frequency effect in their second language than in their first.

In addition to the above findings, L2 learners who reported greater exposure to and (spoken) use of English recognized English words more quickly overall, indicating that exposure to and use of English may facilitate L2 word recognition (even when semantic priming is not observed). Of particular interest here is the indication that automaticity of L2 lexical recognition was impacted by exposure to English and hours spent speaking English per day even when time was controlled. However, because *use* was calculated based on subjects' response to a single question in the Language Experience Survey (see Chapter 3 and Appendices B and C), this finding should be interpreted with some caution. Learners who rated themselves as more proficient recognized target words more slowly, suggesting that self-assessments of proficiency may not be accurate reflections of L2 learners' lexical knowledge, at least to the degree that such knowledge impacts the speed with which they are able to accurately recognize words in English.

Time (of data collection) also had a significant effect on L2 learners' overall RTs (Table 5.10). Learners were slowest to accurately recognize real words at Time 2, faster at Time 3, and fastest at Times 1 and 4. If we were to ignore Time 1, the data might be indicating a practice effect. However, given that Time 1 was one of the fastest data collections (along with Time 4), the relationship between time and overall L2 RTs remains inconclusive. Finally, the fact that self-assessments of language proficiency have been known to be inaccurate is one that has been attested by previous research (e..g, Gardner et al., 1987; Ready-Morfitt, 1991).

This study had a key limitation concerning the method of data collection. Unlike previous studies investigating semantic priming, subjects in the current study participated in the experiment online from an unknown location and an unknown device. Even though a two-step outlier removal process was conducted and inaccurate observations were removed prior to analysis, it is possible that data obtained in a lab, where participant behavior could be better controlled, might have resulted in different results. A future investigation of these data could use these stimuli with L2 learners in a more traditional lab setting to determine whether or not semantic priming is observed when more environmental factors are controlled.

It should also be acknowledged that the division of prime-target word pairs into the sub-categories subsumed within the "related" condition (Appendices E and F) is not an exact science. While stimuli were reviewed independently by three researchers to ensure agreement, it is not always possible, for example, to claim that certain word pairs which lexically co-occur have zero conceptual relationship to one another (e.g., *aluminum-can*). As such, the permeability of category boundaries, especially between position-based and meaning-based associations is a potential limitation. An implication from the results is that form-based associations (e.g., *warn-cold*) should be analyzed separately from position-/meaning-based associations since L1 and NES data suggested that form-based prime-target relationships may interfere with word recognition.

Another implication is methodological. The current study controlled for multiple variables known to influence lexical decision behavior. These included individual differences (e.g., exposure to English), linguistic features (e.g., frequency), and procedural variables (e.g., trial order, stimuli sublist). Despite the fact that exposure to English, target word frequency, trial order, and stimuli sublist all explained significant variance in L2 lexical decisions (see Table 5.9), many of these factors have not been included in previous research, particularly in L2 studies. For example, in Frenck-Mestre and Prince's (1997) seminal L2 priming study, the authors developed stimuli that were balanced for frequency, but this variable was not included in statistical analysis; nor were procedural variables such as trial order included. In another well-cited study, Devitto and Burgess (2004) analyzed L2 participants' initial age of exposure to English and their vocabulary, but it is not apparent that any other individual differences or procedural variables were controlled for statistically in the analysis of their results. Future studies would also do well to continue including a range of control variables related to demographic differences, individual differences, linguistic features of stimuli, and procedural variables, as the findings of the current study suggest that these may explain variation in lexical decision behavior that could otherwise be interpreted as evidence of semantic priming.

5.3 Summary

This chapter was motivated by three key questions. The first was whether L2 learners demonstrated evidence of increased L2 semantic priming over time (2a). Results indicated that L2 subjects demonstrated no evidence of L2 semantic priming at any point in the study. Furthermore, there appeared to be no differential impact on the prime-target relationship and the likelihood of L2 semantic priming taking place over time (2b), as there was no significant interaction between time and the relatedness condition in the L2 learner linear-mixed effects model. However, a trend was observed whereby L2 learners' responses to prime-target conditions ran parallel to NES subjects', with meaning-based associations resulting in the quickest reaction times and formally-associated cues and targets obstructing (rather than facilitating) word recognition. This trend suggests that L2 learners may be responding to related prime-targets in ways that are qualitatively similar to NES subjects but have not yet developed the automaticity for these similarities to emerge quantitatively (i.e., statistically).

Finally, because no priming behavior was observed in L2 learners there were no insights gained regarding the relation between L2 learners' longitudinal priming behavior and vocabulary size (3b). It should be noted, however, that vocabulary size (as reflected by VLT scores), remained unchanged among L2 learners over the course of data collection and was not a significant predictor of overall reaction times (regardless of the relatedness condition) in L2 learner or UG subjects.

Study 2 findings also aligned with what might be predicted from a (non-priming) lexical decision task: Both L2 learners' and NES subjects' reaction times were positively facilitated by the English target words' frequency (with higher frequency words recognized faster) and their trial order, while FB subjects' reaction times were also significantly impacted by trial order. L2 learners' lexical decision behavior was further influenced by their exposure to English and the average number of hours spent speaking English per day. However, L2 learners who self-reported greater proficiency in English actually responded more slowly in the lexical decision task than those whose self-reported proficiency was lower.

5.4 Conclusion

This chapter investigated the development of L2 semantic priming over time in adult English language learners' lexical decision behavior. Results found no evidence of L2 semantic priming (Research Questions 1b and 2b), while some evidence of semantic priming was observed in NES and FB data collected for baseline comparison. Even though statistical evidence of priming was not observed in L2 learner data, learners' reaction times across prime-target conditions ran parallel to NES subjects', with meaning-based associations resulting in the quickest reaction times and formally-associated prime-targets seeming to slow down word recognition. Neither L2 nor UG lexical decision behavior was impacted by vocabulary size (Research Question 3b). On the whole, results suggest that these L2 learners had not yet developed automaticity in L2 word recognition to the degree that semantic priming effects could be observed in their lexical decision times.

Additional insights regarding L2 lexical proficiency may be gleaned from these data: 1) The speed with which L2 learners accurately recognize words in English is impacted more by exposure to English and daily time spent using English than it is by overall time in an English-speaking context; 2) L2 learners demonstrated significant frequency effects for (non-primed) L2 word recognition even at early stages of data collection; and 4) self-reported proficiency may be an unreliable indicator of lexical proficiency as operationalized by automaticity of lexical access.

6 LEXICAL OUTPUT: STUDY 3

While computational indices have proven successful at explaining some variation in human ratings of lexical proficiency and lexical growth over time, applied natural language processing (NLP) approaches are often limited to the analysis of the linguistic and psycholinguistic features of individual words. This approach does little to shed light on the manner in which those lexical items are understood in relation to other items in the mental lexicon. One solution is to utilize computational indices that measure and assess the relationship of individual words to other words in the lexicon. For example, some researchers have begun to use NLP indices to investigate the production of collocations and multiword units (e.g., Crossley & Salsbury, 2011) or the association between lexical items and the more schematic lexicogrammatical constructions in which they frequently occur (e.g., Kyle, 2016).

The goal of Study 3 (S3) was to assess the status of lexical network knowledge, as well as potential growth in L2 lexical knowledge, using a collection of NLP indices intended to measure unique aspects of network knowledge. S3 involved the analysis of a corpus of spoken narratives prompted via video elicitation and recorded using online survey software. During the experiment, adult English language learners were instructed to watch a short animated film with no dialogue and then immediately retell what happened in the film—from memory—in three minutes or less. L2 learners were asked to participate in S3 (watching a different video each time) once every two months for six months.

Study 3 was motivated by the following questions:

1c. What evidence is there of growth in lexical network knowledge as demonstrated by NLP indices capturing associational knowledge measured in L2 learners' spoken lexical production?

2c. Do the above measures develop as a function of time and/or vocabulary size?

6.1 Method

6.1.1 Participants

S3 analyzed the spoken lexical production of L2 learners (*n* = 21) longitudinally. All L2 learners participated in data collection once every two months for up to six months (four times total). Chapter 3 offers further details regarding the recruitment and background of all S3 participants.

6.1.2 Stimuli

Eight 3-minute animated films containing background sound but no dialogue were selected for narrative elicitation purposes. Videos were counterbalanced so that subjects encountered the videos in a different order over the course of the study. See Appendix G for a complete list of the videos used in Study 3 and their references.

6.1.3 Procedure

Subjects completed S3 via the online experiment platform, Testable (www.testable.org). Upon beginning S3, subjects were told that they would watch a short, silent film and be asked to call a phone number and immediately re-tell what they had just see in the film. Once the experiment began, subjects would not be able to pause or re-start the film. When the film ended, subjects were prompted to call the phone number listed on their screen. This phone number automatically linked to a Google Voice account where subjects were prompted to orally re-tell what they had just seen in the film. Subjects could hang up when they were finished speaking. However, because Google Voice only records voicemail for up to three minutes, any subjects who was still speaking after three minutes was cut off.

All trials in S3 were preceded by a practice video. Subjects participated in S3 four times (once every two months) and saw one practice video and one trial video per data collection. In order to increase the amount of language production available for observation in the current study, practice videos were included in the analysis below.

6.1.4 Analysis

6.1.4.1 Transcription

Recorded speech was transcribed by the author and a team of hired transcribers. For details regarding transcription conventions followed by all transcribers, see Appendix H. Any verbal (e.g., *um*) and non-verbal fillers (e.g., [cough]) were removed from the corpus prior to analysis.

6.1.4.2 NLP measures

Automatic text analysis tools—TAALES (Kyle & Crossley, 2015) and TAASSC (Kyle, 2016)—were used to obtain five measures calculated to assess the associative properties of both single- and multi-word[18] lexical items. Each was selected to operationalize a different aspect of subjects' lexical knowledge in relation to other items in their mental lexicon. First, two measures were selected to operationalize the distinctiveness of the lexical and semantic contexts in which a produced word typically occurs (Berger, Crossley, & Kyle, 2017). Next, two measures were chosen to assess the association strength of bigrams (e.g., *take over*) and trigrams (e.g., *on the way*) produced. A fifth measure was included to capture the knowledge of verbs in lexicogrammatical constructions (Römer, O'Donnell, & Ellis, 2014), specifically verb-argument

---

[18] While there are a variety of ways to operationalize and investigate multiple-word phraseological units (e.g., Biber & Conrad, 1999; Granger & Meunier, 2008; Siyanova & Schmitt, 2008; Wray, 2002), the specific approach taken here involves identifying continuous words which recur in a corpus. As such, the current approach fails to take into account formulaic language with variable slots (e.g., Moon, 1997), discontinuous collocations (e.g., Evert, 2008), phraseological preferences (e.g., Cowie, 1998), etc.

constructions (or VACs). The five measures analyzed are listed in Table 6.1 and described in more detail below. When available type-based counts (vs. token-based) were preferred to ensure that the repetition of lexical items did not artificially increase measures.

Table 6.1 Study 3 Associational Measures Selected as Dependent Variables

Measure	**Description**	**Subconstruct**	**Reference**	**Types or Tokens**	**Tool**
University of South Florida association stimuli counts (USF)	Semantic variability of contexts (1,000-word chunks of text) in which word occurs; based on natural log of mean LSA cosine; content words only	Contextual distinctiveness	Nelson, McEvoy, & Schreiber (1998)	Types	TAALES (Kyle & Crossley, 2015)
Semantic distinctiveness (SemD)	Number of different stimuli that elicit word as response in free association; content words only	Contextual distinctiveness	Hoffman, Ralph, and Rogers (2013)	Types	TAALES (Kyle & Crossley, 2015)
COCA spoken bigram association strength	Reports average association strength between bigrams in text with MI^2 scores	Ngram association strength	Davies, 2009	Types	TAALES (Kyle & Crossley, 2015)
COCA spoken trigram association strength	Reports average association strength between bigrams to unigrams in trigrams	Ngram association strength	Davies, 2009	Types	TAALES (Kyle & Crossley, 2015)
Verb-VAC frequency	Production of more frequent main verb lemma–VAC combinations	Lexicogrammatical knowledge	Kyle, K. (in press)	Tokens	TAASC (Kyle & Crossley, 2015)

The TAALES University of South Florida (USF) association norms measure reports the number of stimuli words that resulted in production of the target word as an associate in a word association task. Association data used to calculate the USF measure was collected by Nelson, McEvoy, and Schreiber (1998) from NES subjects in response to 5,019 stimulus words, resulting in a total of 10,470 response words. Words that rank high in the USF measure are associated with a greater variety of words and are thus more likely to come to mind in response to a variety of cues (see Nelson et al., 1998 for details). For example, the word *love* was produced in response to 181 different stimuli and has more readily available associations (and is thus less contextually distinct) than a word like *bride*, which was produced in response to only 6 stimuli in the USF WA task.

The TAALES semantic diversity (SemD) measure operationalizes a word's semantic diversity (or ambiguity) based on the variability of semantic contexts in which it occurs. Derived from Latent Semantic Analysis (Landauer et al., 1998; LSA), SemD was originally calculated by Hoffman, Ralph, and Rogers' (2013) analysis of 1,000-word "contexts" in the written BNC (2007). The measure includes values for 31,739 English words. A high SemD value indicates that a word is more contextually variable, or diverse (i.e., occurring in a variety of semantic contexts), than a lower SemD value. For example, the word *time* has a SemD value of 2.30 and is thus more semantically diverse (or ambiguous) than the word *puppy*, which has a SemD value of 0.93. The assumption motivating inclusion of SemD in the current study is that words with higher SemD values have more semantic associations than words with lower SemD values.

While frequency-based indices of multi-word units are a common measure of lexical knowledge in learner output, corpus and psycholinguistic researchers have begun to explore association-based measures for their potential insights into the cognitive status of multi-word units in native-speakers' minds (Gries, 20015; Ellis & Gries, 2015; Pecina, 2009). Two such association measures were selected for the current study: These report the number of bigrams and trigrams, respectively, in a text and their corpus-derived statistics that quantify their form-function contingency (i.e., the strength of association between words) (cf. Pecina, 2009) in the spoken version of COCA (Davies, 2009). The two measures analyzed were selected, in part, because of their relatively weak correlation with one another ($r = .06$; Table 6.3). Specifically, a TAALES bigram measure *(*spoken COCA, MI^2) was selected to report the number of bigrams in a text based on mutual information (MI^2; Manning & Schütze, 1999), and a TAALES trigram measure (spoken COCA, 2_ΔP) was selected to report the number of strongly associated trigrams in a text based on *ΔP* scores (bigram to trigram) (Ellis, 2007). The mutual information measure can be interpreted as calculating the strength of association between two words, while the *ΔP* score (a hypothesis test) can be thought of us calculating the confidence with which there is an association between words (Clear, 1993; Schmitt, 2010). These two approaches to quantifying contingency were intentionally selected to include both approaches to measuring psychologically associated multiword units. The directionality of the particular *ΔP* trigram measure (i.e., bigram to unigram) selected was preferred due to its weak correlation with bigram measure (Table 6.3).

A second tool (TAASSC; Kyle, 2016) was used to obtain a measure assessing the degree of association between a given verb and the verb-argument construction with which it commonly associates in general English usage. The motivation behind inclusion of this measure was the assumption that verb-argument patterns are an important meaning representation in knowledge of verbs (Römer et al., 2014) and thus constitute another dimension of lexical knowledge. Specifically, a TAASC verb-VAC frequency component score measure was selected for analysis. This measure was derived from a principle component analysis whereby TAASC indices were reduced to a smaller number of components, each representing a group of co-occurring, and thus related, features. The verb-VAC frequency component reports a higher score for production of more frequent main verb lemma–VAC combinations. The measure may also increase with production of more frequent adjective complements and nominal complements. Unlike the bigram and trigram measures described above, which are based on spoken sections of COCA, the verb-VAC frequency measure analyzed here compares VACs production in spoken output to the written sections of COCA, which may be an unsubstantiated comparison. In addition, the VAC index relies on a parser (the Stanford Neural Network Dependency Parser; Chen & Manning, 2014). Parsers are generally unreliable with spoken data (Caines & Buttery, 2014; Hayes et al., 1986; Menzel; 2010) and are likely even more unreliable on L2 spoken data. Thus, there are major limitations with using the verb-VAC frequency measure selected here. It was selected for analysis as an exploratory variable that may have strong construct validity with the knowledge that more reliable verb-argument construction NLP measures do not exist.

6.1.4.3 Statistical analysis

In order to examine the development of associational knowledge in L2 learners' lexical output, five separate linear mixed effects models were conducted to predict the above measures obtained by the text analysis tools. Fixed effects included time (Data collection 1, 2, etc.), English exposure, average number of hours per day spent using English, vocabulary size, and demographic information. Random effects included item (i.e., which of the 8 videos was shown; see Appendix G) and subject.

6.2 Results and discussion

6.2.1 Descriptive Statistics

The initial spoken L2 learner corpus contained 159 texts, including responses to practice videos. Prior to analysis, 47 of these texts were removed because they contained less than 50 words and were too short to produce an adequate enough linguistic signal for interpretation. The resulting corpus contained 112 texts and 14,037 words in total[19]. Table 6.2 displays descriptive statistics for the dependent measures derived from the spoken corpus, as well as for word count. Prior to analysis, all dependent measures were checked for normality and multicollinearity ($r < .70$). To stabilize the variance of the verb-VAC frequency measure, a log10(Y + 10) transformation was performed. Table 6.3 shows that there was no multicollinearity ($r > 0.70$) among dependent measures.

6.2.2 USF stimuli counts LME

An LME model predicting USF stimuli counts in the spoken production of L2 learners found no significant difference in USF stimuli for experimental variables, though one control condition (age) did demonstrate significant differences (Table 6.4). Post-hoc contrast analysis found no significant differences for time when the model was re-leveled to other time periods for comparison. Table 6.4 displays the coefficients, standard error, *t* values, and *p* values for this model.

Significant ($p < 0.05$) and approaching significant ($p < 0.10$) variables from the model in Table 6.4 were then entered into a final USF LME model reported in Table 6.5. This final model reported a marginal R^2 of 0.081 and a conditional R^2 of 0.509. Visual inspection suggested the

---

[19] Due to word length minimums, only 20 subjects' production was ultimately analyzed.

model was not impacted by heteroscedasticity. Table 6.5 displays the coefficients, standard error, *t* values, and *p* values for the model.

Table 6.2 Study 3 Lexical Output Measures, Coverage and Descriptive Statistics

Measures	Word Count		USF stimuli count		Semantic diversity		Bigram association strength		Trigram association strength		Verb-Vac Frequency*	
Coverage	n/a	n/a	0.583	0.059	0.602	0.056	0.832	0.060	0.379	0.090	unknown	unknown
Descrip.	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Overall	125.330	63.100	47.321	10.698	1.945	0.053	9.000	0.386	0.102	0.035	0.997	0.199
Time 1	124.833	64.325	47.919	11.012	1.935	0.064	9.023	0.386	0.106	0.0301	1.014	0.218
Time 2	122.313	67.222	47.322	11.606	1.949	0.053	8.923	0.480	0.095	0.0296	1.036	0.185
Time 3	126.296	61.308	46.488	11.462	1.945	0.049	9.036	0.353	0.105	0.0463	0.958	0.201
Time 4	129.044	61.614	47.518	8.432	1.953	0.044	9.034	0.267	0.105	0.0347	0.965	0.189

** Transformed for normality = log10(Y+10); no coverage statistics provided by TAASC*

Note that all measures report type counts with the exception of Verb-VAC frequency

Table 6.3 Correlation Matrix of Study 3 Dependent Variables

	USF stimuli count	Semantic diversity	Bigram association strength	Trigram association strength	Verb-VAC Frequency*
USF stimuli count	1	0.01	0.11	-0.19	0.12
Semantic diversity	0.01	1	0.01	0.07	-0.12
Bigram association strength	0.11	0.01	1	0.06	0.21
Trigram association strength	-0.19	0.07	0.06	1	-0.27
Verb-VAC frequency*	0.12	-0.12	0.21	-0.27	1

** Transformed for normality = log10(Y+10)*

Note that all measures report type counts with the exception of Verb-VAC frequency

Table 6.4 Study 3 Linear Mixed Effects Model Predicting USF Stimuli Counts

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**71.370**	**9.975**	**7.155**	**< .001**
Demographic Variables				
Age*	**-0.416**	**0.174**	**-2.387**	**0.032**
Gender (baseline: male)	-2.891	3.290	-0.879	0.395
Individual Differences				
Hours per day	-0.259	0.755	-0.342	0.733
L1 distance	-0.049	3.089	-0.016	0.988
Self-reported proficiency	-0.583	1.291	-0.452	0.653
Exposure	-0.107	0.259	-0.414	0.680
VLT	-0.052	0.102	-0.507	0.614
Procedural Variable				
Trial order (baseline: Trial 1)	-5.479	4.509	-1.215	0.271
Experimental Variables				
Time (baseline: Time 1)				
Time2	-2.068	2.125	-0.973	0.333
Time3	-2.550	2.248	-1.134	0.260
Time4	-0.955	2.423	-0.394	0.694

**Significant predictor of reaction times.*

Table 6.5 Study 3 Linear Mixed Effects Model Predicting USF Stimuli Counts, Significant Variables Only

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**59.558**	**4.895**	**12.167**	**<0.001**
Demographic Variables				
Age*	**-0.409**	**0.144**	**-2.837**	**0.012**

**Significant predictor of reaction times.*

Overall, older subjects produced words with a lower USF stimuli count. Words ranking low in the USF measure are produced in response to fewer stimuli in a word association task and are thus more contextually distinct, while words ranking high in the USF measure are associated with a wider variety of words (Nelson et al., 1998).

6.2.3 SemD LME

An LME model predicting semantic diversity in the spoken production of IEP subjects found no significant difference in semantic diversity for experimental or control variables (Table 6.6). This model reported a marginal R^2 of 0.067 and a conditional R^2 of 0.327. Visual inspection suggested the model was not impacted by heteroscedasticity. Table 6.6 displays the coefficients, standard error, *t* values, and *p* values for the L2 model. Post-hoc contrast analysis found no significant differences for time when the model was re-leveled to other time periods for comparison.

6.2.4 Bigrams in Spoken COCA (MI^2)

An LME model predicting the production of statistically associated (MI^2) bigrams (based on spoken COCA) found no significant difference in production of frequent bigrams for experimental or control variables (Table 6.7). Trial order, which approached significance ($p <$ 0.10), was entered as a single fixed effect into a second model but still did not reach significance (that additional model is not included).

Table 6.7 displays the coefficients, standard error, *t* values, and *p* values for the model predicting bigrams. This model reported a marginal R^2 of 0.088 and a conditional R^2 of 0.518. Visual inspection suggested the model was not impacted by heteroscedasticity. Post-hoc contrast analysis found no significant differences for time when the model was re-leveled to other time periods for comparison.

Table 6.6 Study 3 Lineare Mixed Effects Models Predicting Semantic Diversity

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**1.946**	**0.042**	**45.792**	**p < 0.001**
Demographic Variables				
Age	0.000	0.001	0.046	0.964
Gender (baseline: male)	0.001	0.013	0.060	0.953
Individual Differences				
Hours per day	0.004	0.004	1.085	0.283
L1 distance	-0.019	0.012	-1.624	0.124
Self-reported proficiency	0.001	0.006	0.224	0.824
Exposure	-0.001	0.001	-0.577	0.568
VLT	0.000	0.000	0.826	0.417
Procedural Variable				
Trial order (baseline: Trial 1)	0.014	0.022	0.655	0.537
Experimental Variables				
Time (baseline: Time 1)				
Time2	0.006	0.012	0.517	0.606
Time3	0.008	0.013	0.599	0.551
Time4	0.014	0.013	1.026	0.307

**Significant predictor of reaction times.*

Table 6.7 Study 3 Linear Mixed Effects Model Predicting Spoken Bigrams in COCA

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**9.297**	**0.454**	**20.482**	**p < 0.001**
Demographic Variables				
Age	-0.002	0.009	-0.193	0.849
Gender (baseline: male)	-0.008	0.164	-0.049	0.962
Individual Differences				
Hours per day	-0.030	0.031	-0.972	0.333
L1 distance	-0.035	0.154	-0.229	0.822
Self-reported proficiency	0.052	0.053	0.966	0.337
Exposure	0.002	0.011	0.211	0.833
VLT	-0.002	0.004	-0.380	0.705

Procedural Variable				
Trial order (baseline: Trial 1)	-0.202	0.098	-2.057	0.090
Experimental Variables				
Time (baseline: Time 1)				
Time2	-0.125	0.083	-1.502	0.137
Time3	-0.018	0.089	-0.204	0.839
Time4	0.045	0.099	0.457	0.648

6.2.5 Trigrams in Spoken COCA

An LME model predicting the production of statistically associated (ΔP) trigrams (based on spoken COCA) found that gender statistically predicted production of frequent trigrams (Table 6.8), with women producing significantly more frequent trigrams than men. Table 6.8 displays the coefficients, standard error, *t* values, and *p* values for this model. Gender, the only significant ($p < 0.05$) variable from the model in Table 6.8, was then re-entered into the final trigram LME model reported in Table 6.9. This model reported a marginal R^2 of 0.190 and a conditional R^2 of 0.426. Post-hoc contrast analysis found no significant differences for time when the model was re-leveled to other time periods for comparison.

Table 6.8 Study 3 Linear Mixed Effects Model Predicting Production of Spoken Trigrams

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**0.128**	**0.033**	**3.863**	**p < 0.001**
Demographic Variables				
Age	-0.001	0.001	-0.947	0.359
Gender (baseline: male)*	**0.024**	**0.011**	**2.170**	**0.047**
Individual Differences				
Hours per day	-0.030	0.031	-0.972	0.333
L1 distance	-0.035	0.154	-0.229	0.822
Self-reported proficiency	0.052	0.053	0.966	0.337
Exposure	0.002	0.011	0.211	0.833
VLT	-0.002	0.004	-0.380	0.705

Procedural Variable				
Trial order (baseline: Trial 1)	0.003	0.010	0.301	0.773
Experimental Variables				
Time (baseline: Time 1)				
Time2	-0.012	0.008	-1.508	0.135
Time3	-0.001	0.008	-0.095	0.925
Time4	-0.005	0.009	-0.559	0.578

**Denotes statistical significance*

Table 6.9 Study 3 Linear Mixed Effects Model Predicting Production of Spoken Trigrams, Significant Variables Only

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**0.083**	**0.010**	**8.536**	**p < 0.001**
Demographic Variables				
Gender (baseline: male)*	**0.024**	**0.010**	**2.343**	**0.031**

**Denotes statistical significance*

6.2.6 Verb-VAC LME

Finally, an LME model predicting verb-VAC frequency for L2 learners found no significant difference in production of verb-VAC frequency for experimental or control variables, though VLT score approached significance (see Table 6.8). Contrary to expectations, subjects who scored lower on the VLT showed a trend toward producing more frequent verb-VAC associations. However, when VLT was entered as a single fixed effect into a second model, it did not reach significance (so that additional model is not included). Table 6.10 displays the coefficients, standard error, *t* values, and *p* values for the model predicting verb-VAC frequency. This model reported a marginal R^2 of 0.100 and a conditional R^2 of 0.226. Visual inspection suggested the model was not impacted by heteroscedasticity. Post-hoc contrast analysis found no significant differences for time when the model was re-leveled to other time periods for comparison.

Table 6.10 Study 3 Linear Mixed Effects Model Predicting Verb-VAC Frequency

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**0.893**	**0.181**	**4.922**	**p < 0.001**
Demographic Variables				
Age	0.002	0.003	0.633	0.535
Gender (baseline: male)	0.003	0.059	0.051	0.960
Individual Differences				
Hours per day	-0.021	0.016	-1.294	0.200
L1 distance	0.083	0.056	1.499	0.153
Self-reported proficiency	0.018	0.027	0.661	0.511
Exposure	0.008	0.005	1.508	0.139
VLT	-0.004	0.002	-1.846	0.074
Procedural Variable				
Trial order (baseline: Trial 1)	-0.034	0.046	-0.739	0.489
Experimental Variables				
Time (baseline: Time 1)				
Time2	0.003	0.049	0.070	0.944
Time3	-0.066	0.051	-1.290	0.200
Time4	-0.032	0.055	-0.595	0.553

** Denotes statistical significance*

^ Approaching statistical significance (p < 0.10)

Note: The verb-VAC frequency measures was transformed (log10(Y+10) prior to analysis

6.2.7 Discussion

Study 3 (S3) analyzed a learner corpus of transcribed spoken narratives produced in response to short animated films with no dialogue. Each subject participated once every two months up to four times each. The goal of S3 was to assess the longitudinal development of L2 lexical network knowledge in adult English learners' spoken output using a collection of NLP indices selected to measure unique aspects of associational lexical knowledge. These included word association stimuli counts (USF), a semantic distinctiveness measure (SemD), bigram and trigram association strength measures, and a verb-VAC frequency score (see Table 6.1). Analysis demonstrated no evidence of statistical change over time in lexical network knowledge as operationalized by the NLP indices listed above (Research Question 1c). Nor was there evidence of a relationship between any of these measures and vocabulary size (Research Question 2c), with the exception of a trend in verb-VAC frequency whereby learners who scored lower on the VLT were more likely to produce frequent verb-VAC combinations (Table 6.10). However, this trend was not statistical ($p > .05$).

Significant main effects across models in S3 included age in the analysis of USF word association stimuli counts (Table 6.5) and gender in the analysis of trigrams (Table 6.9). Across all time periods, older subjects produced words with a lower USF stimuli count. In other words, older L2 learners were more likely to produce words associated with fewer stimuli in a large-scale NES word association task (Nelson et al., 1998). It is not immediately clear why this would be the case. However, a previous study found that USF correlated negatively with lexical proficiency, meaning that speakers who were rated as more lexically proficient used fewer words elicited by a range of stimulus words in the USF (Berger, Crossley, & Kyle, 2017). It may be that older L2 learners who contributed to this spoken corpus were on the whole more lexically proficient. Indeed, the correlation between age and VLT score in L2 participants indicated a positive, though weak, effect size ($r = 0.154$, $p > 0.416$). In addition to the main effect of age, L2 learners who were women produced more frequent trigrams in their speech than men.

While it is not immediately clear why the dependent variables analyzed in the current study showed no growth over time, it is unlikely that these L2 learners, all of whom were students in Intensive English Programs (IEPs) at American universities, simply failed to develop spoken lexical knowledge in the six months during which this study took place. It should be acknowledged that the dependent measures selected for analysis each operationalize one specific aspect of associational lexical knowledge. Other lexical and psycholinguistic characteristics of spoken L2 language that have demonstrated longitudinal growth in previous research include measures such as concreteness, hypernymy, polysemy, range, imageability, familiarity, and frequency (Crossley & McNamara, 2013; Crossley, Salsbury, & McNamara, 2009, 2010; Crossley, Salsbury, & McNamara, 2011, 2013; Crossley, Salsbury, McNamara, & Jarvis, 2010; Crossley, Salsbury, McNamara, & Jarvis, 2011; Kyle & Crossley, 2015). However, these measures were not included in the current study, as lexical network knowledge was of primary interest to this dissertation project and its research questions.

It should also be acknowledged that the current study took just one approach to operationalizing production of multi-word units (i.e., continuous bigrams and trigrams), but there are a number of other ways to investigate and identify knowledge of multiple-word phraseological units (e.g., Biber & Conrad, 1999; Granger & Meunier, 2008; Siyanova & Schmitt, 2008; Wray, 2002). While the production of associated bigrams and trigrams did not demonstrate growth over time, it may be that these same learners were developing lexical knowledge with regard to other multi-word phenomena, such as idioms (Moon, 1997), semantic prosody (Hunston, 2007), collocations (Evert, 2008), or lexical phrases with variable slots (Gray & Biber, 2013). It is also entirely possible that the development that IEP students who participated in the current study experienced with regard to their spoken language over the course of six months was not exclusively lexical. In other words, they may have shown improvement in any number of constructs that were not assessed in the current study, such as fluency, prosody, pronunciation, pragmatic ability, etc. While these certainly involve lexical knowledge, they rely on resources beyond those analyzed here.

Another potential explanation for the lack of significant findings in the current study is methodological. With the exception of the verb-VAC frequency component score, all measures analyzed as dependent variables in the current study reported types (vs. tokens). This is a departure from some previous applied NLP studies which have relied exclusively on indices that report tokens (e.g., Berger, Crossley, & Kyle, 2017a, 2017b; Kyle & Crossley, 2015; Salsbury, T., Crossley, S. A., & McNamara, D. S., 2011), though others have examined both (e.g., Crossley et al., in press; Crossley et al., 2014; Crossley, Kyle, & Salsbury, 2016). Because the former approach allows for the repetition of a word(s) within a single text to increase the value of the measures assessed, it could be argued that the scores reported conflate repetition and/or fluency (i.e., the production of more words) with lexical development (i.e., the availability of specific words within one's mental lexicon)[20]. In this regard, the current approach could ultimately reflect a more accurate representation of what is available in learners' lexicons at any given point in time. It could also mean that measures which have been indicative of lexical proficiency or development in previous studies fail to show evidence of the same in the current study.

To further investigate the above, a posthoc analysis of token-based (vs. type-based) measures was assessed on Study 4 data. Table 6.11 compares both the type-based stimuli count measures (used in analyses above) with equivalent token-based measures (as have typically been used in studies of this nature). While a greater range of variability around the mean can be observed for all token-based measures, there were no statistically significant differences across time for either type-based measures or token-based measures (Table 6.11). While it is reasonable to argue that type-based measures reflect a more accurate representation of learners' mental

---

[20] Such repetition may be of interest to researchers investigating other constructs (e.g., lexical diversity, fluency, lexical representation, lexical sophistication). Thus the usefulness of type vs. token counts is relative to a one's research questions and a matter of construct validity.

lexicon, the difference between type-based and token-based measures does not account for the discrepancy in findings between the current study and previous research.

Table 6.11 Study 3 Posthoc Analysis of Type vs. Token Counts for TAALES Measures

Measures	**USF types**		**USF tokens**		**SemD types**		**SemD tokens**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Overall	47.321	10.698	47.927	16.551	1.945	0.053	1.896	0.059
Time 1	47.919	11.012	48.947	18.175	1.935	0.064	1.935	0.064
Time 2	47.322	11.606	48.936	18.004	1.949	0.053	1.949	0.053
Time 3	46.488	11.462	46.102	14.693	1.945	0.049	1.945	0.049
Time 4	47.518	8.432	47.309	15.010	1.952	0.045	1.952	0.045
	F	*p*	*F*	*p*	*F*	*p*	*F*	*p*
	0.087	0.967	0.193	0.901	0.568	0.637	0.896	0.446
	Bigram Types		**Bigram tokens**		**Trigram types**		**Trigram tokens**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Overall	9.000	0.388	9.189	0.439	0.102	0.035	0.103	0.036
Time 1	9.023	0.386	9.285	0.387	0.106	0.030	0.108	0.033
Time 2	8.923	0.480	9.062	0.545	0.095	0.030	0.095	0.032
Time 3	9.036	0.353	9.228	0.411	0.105	0.046	0.106	0.046
Time 4	9.037	0.273	9.194	0.339	0.103	0.034	0.104	0.034
	F	*p*	*F*	*p*	*F*	*p*	*F*	*p*
	0.590	0.623	1.463	0.229	0.655	0.582	0.760	0.519

* *Indicates significance (p < .05)*

Still, the results of the post-hoc analysis behoove researchers conducting similar research in the future to differentiate between type and token counts when statistically analyzing NLP indices, as relying exclusively on the latter may inaccurately conflate fluency with lexical knowledge. Of course, the decision of whether to analyze type- or token-based indices depends entirely upon which is a more valid operationalization of the particular construct under investigation.

Because dependent variables of interest failed to show development over time as anticipated (and because the measures investigated here arguably assess a very specific aspect of lexical knowledge), an additional post-hoc analysis was conducted to determine whether a more classic lexical measure, concreteness (Paivio, 1971, 2013), might indicate growth over time in these data, as it has been demonstrated to do in other L2 longitudinal spoken datasets (Crossley & Skalicky, in press; Salsbury et al., 2011). An LME (Table 6.12) found no significant difference in production of concreteness[21] over time and post-hoc contrast analysis found no significant differences for time when the model was re-leveled to other time periods for comparison. The model in Table 6.12 reported a marginal R^2 of 0.076 and a conditional R^2 of 0.550. Visual inspection suggested the model was not impacted by heteroscedasticity.

Table 6.12 Study 3 Linear Mixed Effects Model Predicting Concreteness

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**3.591**	**0.270**	**13.305**	**0.000**
Demographic Variables				
Age	0.002	0.005	0.497	0.626
Gender: male	-0.116	0.090	-1.283	0.218
Individual Differences				
Hours per day	-0.005	0.020	-0.243	0.808
L1 distance	-0.015	0.085	-0.171	0.866
Self-reported proficiency	0.036	0.034	1.036	0.303
Use	-0.012	0.007	-1.820	0.073
VLT	-0.002	0.003	-0.775	0.442
Procedural Variable				
Trial order (baseline: Trial 1)	0.068	0.119	0.569	0.591
Experimental Variables				
Time (baseline: Time 1)				
Time2	0.076	0.055	1.378	0.172
Time3	0.044	0.059	0.753	0.454

[21] For the concreteness dependent variable, a TAALES (Kyle & Crossley, 2015) measure based on mean concreteness (Brysbaert et al., 2013) was assessed.

Time4	0.000	0.064	-0.006	0.995

** Denotes statistical significance*

The results of this post-hoc analysis suggest that even a classic index of lexical development, such as concreteness, failed to demonstrate growth in L2 learners' spoken production over the course of data collection. As recommended previously, future work with this same dataset could investigate whether non-lexical spoken skills (such as fluency, prosody, etc.) may have been more likely to show evidence of growth.

A key limitation of the current study is the unnatural manner in which spoken language was elicited. Subjects were prompted to call a phone number and speak their video re-telling aloud into a voicemail recording. Given the fact that there was no actual human interlocutor present, the language elicited is arguably less natural (and certainly less interactive) than if learners had been recorded using English for more authentic purposes. Future research of this nature would do well to elicit spoken language in more natural contexts, so as to capture more authentic, interactive samples of language usage across a range of topics. It could also be argued that the eight videos selected for the current study portrayed relatively straightforward, linear, and concrete narrative elements, all of which could be conveyed without the need for terribly sophisticated lexical resources. For this reason, future work using video elicitation should also consider the content of the video stimuli and the extent to which a range of lexical production is elicited.

In order to determine how well the language elicited by individual videos was consistent in the current study, post-hoc analyses were conducting with video as a fixed (vs. random) effect to examine if the videos were significant predictors of lexical output. No differences were found in terms of the significant variables that predicted USF, SemD, or bigrams. There were changes in significant main effects when predicting trigrams and verb-VAC frequency; however, time remained insignificant in both models. The final tables from this post-hoc analysis (reporting significant variables only) are described below.

Whereas video as a random effect resulted in gender as the only significant main effect for trigram production (Table 6.9), Table 6.13 demonstrates that L1 distance, trial, and video all significantly impacted production of frequent trigrams when video was included in the model as a fixed effect. In addition to women producing more frequent trigrams, learners whose L1 was further from English produced less frequent trigrams over all. Learners were also more apt to produce trigrams in the second video trial during each data collection. Finally, results indicate that some videos resulted in variable production of frequent trigrams, indicating a "prompt" effect.

Table 6.13 Study 3 Posthoc Linear Mixed Effects Model Predicting Production of Spoken Trigrams Using Video as a Fixed Effect, Significant Variables Only

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**0.064**	**0.025**	**2.575**	**0.013**
Demographic Variables				
Gender (baseline: male)*	**0.024**	**0.009**	**2.748**	**0.013**
Individual Differences				
L1 distance*	**-0.023**	**0.008**	**-2.718**	**0.014**
Procedural Variable				
Trial order (baseline: Trial 1)*	**0.037**	**0.010**	**3.533**	**0.001**
Video (baseline: Aviator)*				
Carrots	-0.006	0.011	-0.582	0.562
Chess*	**0.026**	**0.012**	**2.094**	**0.039**
Homeless	-0.019	0.011	-1.775	0.079
Lucky*	**0.041**	**0.011**	**3.575**	**0.001**
Octopus*	**0.039**	**0.011**	**3.610**	**0.000**
Petshop	-0.002	0.011	-0.174	0.862

** Denotes statistical significance (*p*<.05)*

While calculating video as a random effect in earlier analysis revealed no significant main effects for verb-VAC frequency (Table 6.10), re-running the same model with video as a fixed effect indicated that both trial order and video were main effects (Table 6.14). Subjects were more likely to produce frequent verb-VAC combos during the second video trial, and again, some videos resulted in variable production of frequent verb-VAC combos.

Table 6.14 Study 3 Posthoc Linear Mixed Effects Model Predicting Verb-VAC Frequency with Video as Fixed Effect, Significant Variables Only

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	**1.285**	**0.103**	**12.476**	**< .001**
Procedural Variable				
Trial order (baseline: Trial 1)*	**-0.190**	**0.064**	**-2.958**	**0.004**
Video (baseline: Aviator)*				
Carrots	0.130	0.067	1.947	0.055
Chess	-0.114	0.077	-1.492	0.139
Homeless*	**0.153**	**0.066**	**2.322**	**0.022**
Lucky	-0.082	0.070	-1.172	0.244
Octopus	-0.127	0.066	-1.927	0.057
Petshop	0.025	0.066	0.376	0.708

** Denotes statistical significance (*p*<.05)*

As discussed previously, an important limitation of the VAC-verb frequency dependent variable is its reliance on the Stanford Neural Network Dependency Parser (Chen & Manning, 2014), which has proven to be unreliable with spoken data in the past (Kyle, in press). It may be that parser's performance on the learner corpus assessed in the current study was equally unreliable, resulting in the inability to detect change in lexicogrammatical knowledge of verb-argument patterns even if it were present in learners' spoken production. On the whole, more NLP indices of this nature are needed which can reliably analyze spoken data, not to mention spoken learner data.

Finally, due to the low coverage of the selected trigram measure (see Table 6.2) (suggesting that few highly associated trigrams were being produced by learners at all), an additional TAALES measure was investigated in post-hoc analysis to determine what proportion of trigrams produced by L2 learners were shared with the most frequent 60,000 trigrams found in spoken COCA (Davies 2009). An LME model predicting the production of this trigram proportion measure found that only trial order statistically predicted production of common trigrams (Table 6.15).

Table 6.15 Study 3 Linear Mixed Effects Model Predicting Trigram Proportion (60K)

Fixed Effect		Coefficient	Std. error	*t*	*p*
(Intercept)*		**0.186**	**0.083**	**2.248**	**0.034**
Demographic Variables					
Age		0.001	0.002	0.587	0.565
Gender: male		-0.002	0.030	-0.076	0.940
Individual Differences					
Hours per day		-0.004	0.006	-0.700	0.486
L1 distance		0.041	0.028	1.450	0.165
Self-reported proficiency		0.019	0.010	1.913	0.059
Use		0.001	0.002	0.593	0.555
VLT		0.000	0.001	-0.059	0.954
Procedural Variable					
Trial order (baseline: Trial 1)		**-0.055**	**0.011**	**-4.861**	**0.000**
Experimental Variables					
Time (baseline: Time 1)					
	Time2	-0.014	0.016	-0.899	0.371
	Time3	-0.028	0.017	-1.697	0.093
	Time4	0.001	0.019	0.050	0.960

** Denotes statistical significance*

^ Approaching statistical significance ($p < 0.10$)

Significant ($p < 0.05$) and approaching significant variables from the model in Table 6.15 were then re-entered into models until only significant variables remained. The resulting final trigram proportion LME is reported in Table 6.16. This model reported a marginal R^2 of 0.124 and a conditional R^2 of 0.436. Trial order remained a significant predictor of RTs, but there were no other significant main effects. While the initial trigram measure analysis (see Tables 6.8 and 6.9) demonstrated no evidence of the production of more highly associated trigrams over time, this post-hoc analysis indicates that very few trigrams common in native speaker spoken samples were being produced by learners at all (over time or otherwise).

Table 6.16 Study 3 Linear Mixed Effects Model Predicting Trigram Proportion Variable, Significant Variables Only
Table 6.15

Study 3 Linear mixed effects model predicting WTF trigram proportion variable, sig. only

Fixed Effect	Coefficient	Std. error	*t*	*p*
(Intercept)*	0.317	0.013	23.760	0.000
Procedural Variable				
Trial order (baseline: Trial 1)*	**-0.055**	**0.012**	**-4.595**	**0.004**

** Denotes statistical significance*

6.3 Summary and Conclusion

Study 3 sought to assess the development of longitudinal L2 lexical knowledge through analysis of a transcribed learner corpus collected via video elicitation. Five NLP indices were selected as dependent variables to measure unique aspects of associative knowledge that may manifest in learners' spoken lexical production. There was no evidence of longitudinal development of lexical network knowledge as demonstrated by the NLP indices analyzed, nor was there any statistical relationship between vocabulary size (as measured by the VLT) and any of the five dependent variables. Among the possible explanations for the current findings were a focus on associational lexical features with the goal of indexing L2 lexical network knowledge; the unnatural and non-interactive manner by which spoken language was elicited; the exclusion of non-lexical aspects of learners' spoken performance (e.g., prosody); the (potentially) limited range of lexical items elicited by the simple narratives portrayed within these video stimuli; and the unreliable performance of the verb-VAC frequency measure specifically on spoken learner data.

7 CONCLUSION AND OUTLOOK

The goal of this dissertation project was to enhance our understanding of the L2 English lexicon by observing the longitudinal development of L2 depth of lexical knowledge in adult English language learners. Another purpose of the project was to fill a significant research gap in the field for more investigations of longitudinal L2 development in general (Ortega & Byrnes, 2008), as well as for longitudinal investigations of both breadth and depth of lexical knowledge (M. Daller, Turlik, & Weir, 2013; Dóczi & Kormos, 2016; Schmitt, 2010). To this end, three longitudinal experimental studies were conducted to track the development of learners' L2 lexical network knowledge over six months of English language study in an English-speaking context. Methods of data collection included a vocabulary size assessment, a word association task, lexical decision semantic priming, and the computational analysis of learners' spoken lexical output. With the exception of the production of less canonical word associations, few developmental trends were observed over time across datasets. Nor was there an observed relationship between learners' English vocabulary size and their performance in the three experimental studies. These findings have implications for our understanding of L2 lexical network knowledge and its rate of development, as well as methodological implications for future investigations of the L2 lexicon. A summary of the outcomes and findings of this project are provided.

7.1 Summary of Findings

A summary of results for the three longitudinal studies comprising this project are organized by research question below. Because the relationship between vocabulary size and dependent measures obtained from the three experimental studies is included in all three research questions, it is worth noting that learners' VLT scores did not demonstrate statistically significant change over the course of data collection (see Chapter 3 and Table 3.7). In other words, learners' breadth of vocabulary knowledge did not increase over time as might have been expected.

Past studies suggest that such a finding is not entirely anomalous. For example, Schmitt and Meara (1997b), who administered the complete *Vocabulary Levels Test VLT* (Nation, 1990) to EFL learners both prior to and following one year of study, found that despite overall growth in recognition knowledge of words, 28% of their subjects demonstrated actual decreases in knowledge over the course of the year. The authors also noted that individual patterns of development among learners showed a great deal of fluctuation and that L2 vocabulary size can be expected to develop in an incremental and non-linear fashion.

In another study assessing vocabulary size in spoken lexical output, Crossley, Salsbury, and McNamara (2010) found that polysemous words in learners' spoken output exhibited a breadth of vocabulary learning curve based on the power law of learning, "where a steep increase at the beginning is followed by a flattening out of the learning curve in later stages" (M. Daller et al., 2013, p. 213). Despite the fact that the L2 learners assessed in the current project had all begun studying English at an IEP program during the semester of initial data collection[22], it could be that they were far enough along in their overall study of English that such flattening was already taking place. It may also be the case that as students enrolled in an IEP program, their learning was more oriented toward developing skills like academic writing and listening than to vocabulary, per se. However, previous research has demonstrated that vocabulary size does demonstrate longitudinal growth among L2 learners (Crossley, Cobb, & McNamara, 2013;

---

[22] Only those learners who began studying English in an IEP program in the spring semester of 2017 and placed into high beginner through high intermediate levels in their respective programs were recruited for participation.

Crossley, Salsbury, & McNamara, 2009, 2010; Dóczi & Kormos, 2016; Milton, 2009; Qian, 1999, 2000; Schmitt, 2010; Zareva, 2005; Zhang & Lou, 2004).

7.1.1 Research Question 1

What patterns of development are revealed by the longitudinal analysis of L2 learners' English word associations with relation to time and vocabulary size?

Study 1 revealed a trend whereby L2 learners' word associations (WA) became less canonical (i.e., less like baseline WAs collected from NES and/or highly proficient English speakers) over time. This trend was observed in two different approaches to assessing canonicity (see Table 4.3): comparison to pre-existing NES norms (i.e., the USF word association norms dataset) and to a bespoke norms list collected from NES and FB undergraduates who more closely aligned with the L2 learners' demographics[23]. However, analyses indicated that only the trend in USF canonicity was statistical, with learners producing significantly less canonical WAs at Time 3 than at Time 1. These differences held even when demographic variables, individual differences, and procedural variables were controlled in analysis. One reason the canonicity measure derived from the USF norms list may have demonstrated significant change over time—whereas the bespoke norms list did not—is because the former represented a much denser dataset of word associations. Regardless, it is striking that a dataset collected from 6,000 NES participants in 1973 and one collected from 37 NES and FB participants in 2017 resulted in relatively similar patterns when used to analyze learner data. This suggests that word association behavior indeed demonstrates a great deal of systematicity among native and proficient speakers, even across multiple decades.

Post-hoc analysis of four learners whose WAs demonstrated the largest decreases in

---

[23] Recall that a post-hoc analysis demonstrated a similar trend in bespoke canonicity even when FB undergraduates' word associations were withheld from calculation of bespoke scoring.

canonicity across data collection (Table 4.8) found that learners' early WAs were primarily meaning-based, with some collocational (i.e., position-based) associations produced as well. By the end of data collection, however, these same learners' meaning-based associations were highly personal (e.g., *lonely—self* , *religion—Muslim*, *pride—my policy*), and thus less canonical. They were also producing more form-based associations and more idiosyncratic—and arguably more creative—responses overall. These findings support previous research suggesting that L2 speakers' word associations may be vague (Sökmen, 1993) and/or influenced by "feelings, attitudes or strong memories" (Fitzpatrick, 2006, p. 140) and that advanced L2 speakers are more apt to produce form-based associations than native speakers (Fitzpatrick, 2006). At the same time, learners' decrease in canonicity over time, despite the fact that response words continued to bear some relation to the cue (i.e., they were not entirely random) may suggest a relationship between gains in lexical proficiency and creative use of the L2, as has been observed in previous research (Bell, 2005; Bell, Skalicky, & Salsbury, 2014; Belz, 2002).

Taken as a whole, the non-canonical responses produced by L2 learners as they developed lexically may indicate that they are behaving more like proficient L2 speakers (their WAs are exhibiting a great deal of variation, they are producing more creative or form-based associations, etc.) but not necessarily exactly like native-speakers on the WA task. As with the quantitative analysis of canonicity, this qualitative analysis suggests that as learners develop lexically, they may begin behaving more like proficient L2 speakers and less like native-speakers.

This finding that L2 learners' WAs became less—and not more—native-like calls into question the assumption that more native-like responses are the most obvious indication of lexical proficiency in word association behavior (Fitzpatrick, 2006). More broadly, researchers (e.g., Ortega, 2016; Klein, 1998) have begun to question the use of native-like targets across the field of SLA, arguing that reliance on (effectively) monolingual competence as a target implies a deficit perspective toward L2 learning, particularly toward late bilingualism (Ortega, 2014). Ortega (2016) has observed that the many references to "native-likeness" in L2 research could actually be interpreted to mean "monolingual-like." Such an approach, she argues, fails to capture the true development of multiple-language competencies across a person's lifespan (Ortega, 2016).

Finally, canonicity score were not influenced by vocabulary size, indicating that the constructs of lexical knowledge assessed by both tasks (the WA task and the vocabulary size instrument) may be distinct. This finding supports the notion that breadth and depth of vocabulary knowledge do not index the same construct—a finding supported by others (e.g., Crossley, Salsbury, & McNamara, 2010; Nurweni & Read, 1999; Qian, 1999, 2000)—and that tasks designed to assess them may require learners to draw on different resources. At the same time, results of Study 1 suggest that it is possible for depth of knowledge to demonstrate evidence of development (as suggested by WA results) while breadth of vocabulary remains sedentary.

7.1.2 Research Question 2

What patterns of development are revealed by the longitudinal analysis of L2 learners' English semantic priming with relation to time and vocabulary size?

Results from Study 2 demonstrated no evidence of semantic priming in L2 English learners' lexical decision behavior at any point during longitudinal data collection. While learner performance on the semantic priming task did align with what might be expected of lexical decision behavior (specifically, frequent words were recognized more quickly and there was an anticipated practice effect for trial order), no significant decrease in word recognition latencies was observed between related primes and targets when compared to unrelated pairs. This was the case even when more fine-grained relationships between prime and target (beyond simply "related") were examined, including form-based, meaning-based, or position-based associations. In NES data, a trend approaching significance ($p < 0.10$) was observed whereby meaning-based related prime-target pairs were recognized more quickly than unrelated pairs, while a negative statistical reverse-priming effect ($p < 0.05$) was found for form-based associations (i.e., form-related prime-target pairs statistically impeded response latencies compared to the unrelated condition). In FB data, a statistical facilitative priming effect was found for position-based associations compared to the unrelated condition.

Despite there being no statistical evidence of semantic priming among L2 learners, their descriptive RTs times ran parallel to NES RTs: Both subject groups responded more slowly to the form-based condition than the unrelated condition, and meaning-based associations resulted in the fastest latencies, followed by position-based associations. One interpretation of this trend suggests that these L2 learners' lexical decision behavior in response to prime-target conditions was qualitatively similar to NES subjects, but they had yet to develop the automaticity necessary for semantic priming effects to be statistically evident. Future work is needed to validate this claim.

Another explanation for the lack of semantic priming effects in L2 data is that subjects did not respond to target words as quickly as they might have in a lab, rendering some of the data less accurate and precise than they would have been in a more controlled lab setting. It is also possible that the reaction times recorded by the online software used for Study 2 (www.testable.org) may not have been as precise as experimental software in a lab. However, because some evidence of priming effects was observed in UG data, it seems likely that it would have been captured in L2 data as well.

When Study 2 data were analyzed as lexical decisions (rather than for evidence of semantic priming), we saw that time (Data collection 1, 2, 3, or 4), as well as learners' exposure to English, use of English[24], and self-reported proficiency were significant predictors of accurate lexical decision latencies for L2 learners. The role of time in determining RTs was largely inconclusive (see Table 5.10) with learners fastest to accurately recognize real words at Times 1 and 4, a little slower at Time 3, and the slowest at Time 2. However, the automaticity of L2 lexical recognition was facilitated by exposure to English and hours spent speaking English, even when variation in latencies due to time was accounted for. This finding will come as no surprise to language instructors: Those students who are actively engaged with the target language are more likely to benefit than those who are in the same classroom or program for the same amount of time but are less engaged. And due to self-reported proficiency predicting lexical decision latencies in the opposite direction as might be expected, there is evidence that self-assessments of language proficiency may not align always align with learners' more automatic skills, such as lexical access and word recognition (Gardner et al., 1987; Ready-Morfitt, 1991). Limitations of these three dependent variables (*proficiency*, *exposure*, and *use*) should be acknowledged, however. Both *proficiency* and *use* were derived from single responses provided by subjects on the Language Experience Survey, and exposure fails to distinguish between receptive exposure and productive use of English in subjects' daily lives (see Appendices B and C). It could thus be argued that these variables only provide a rough self-estimation of learners' actual proficiency and engagement with their L2.

[24] See Appendix C for an explanation of how exposure and use were derived from learners' responses to the Language Experience Survey

Finally, because vocabulary size (as reflected by VLT scores) was not a significant predictor of overall L2 or UG reaction times (regardless of the relatedness condition), the relationship between vocabulary size and semantic priming (among these particular subjects) remains undetermined. However, as with Study 1, this finding suggests that the breadth of lexical knowledge measured by the VLT may be distinct from the resources learners draw on during L2 lexical recognition.

7.1.3 Research Question 3

Study 3: What patterns of development are revealed by the longitudinal analysis of associational lexical knowledge in L2 learners' spoken production with relation to time and vocabulary size?

Results from Study 3 found no evidence of development of L2 lexical network knowledge in learners' longitudinal output as demonstrated by the five indices analyzed as dependent variables, and there was no statistical relationship observed between vocabulary size (as measured by the VLT) and any of the dependent variables. Potential explanations for the lack of observed change over time include the study's focus on associational lexical features only (with the goal of indexing L2 lexical network knowledge); the potentially unnatural manner in which spoken language was elicited; the exclusion of non-lexical aspects of learners' spoken performance (e.g., fluency, prosoedy, pronunciation); the fact that the simple narratives portrayed in video stimuli may have elicited a limited range of lexical items; and the unreliability of the verb-VAC frequency measure specifically on spoken learner data. Many of the above constitute limitations of the study that should be acknowledged.

A key limitation of Study 3 was the manner in which spoken language was elicited and recorded. Ideally, a human interlocutor would have been present to provide a more natural context while also eliciting interactive elements of speech. The content of the video stimuli used in Study 3 may also have constrained the extent of lexical production they elicited. Because a post-hoc analysis of learners' lexical output found no evidence of development over time in concreteness, a well-attested measure of growth in spoken L2 longitudinal studies (Crossley & Skalicky, in press; Salsbury, et al. 2011), it seems likely that video stimuli played a role in limiting the language produced. Additionally, computational measures such as the verb-VAC frequency measure used in Study 3 are likely to be less accurate with spoken learner data.

It was also hypothesized that the use of type-based (vs. token-based) NLP measures in the current analyses may have been to explain for discrepancies between these findings and others' (e.g., Berger, Crossley, & Kyle, 2017a). While a post-hoc analysis comparing the type-based measures (used in the current analysis) to equivalent token-based measures found that the latter indeed resulted in more variance, that variance did not appear to impact the findings of the current study (see Table 6.11 and Chapter 6 discussion). Future work is needed to determine the role of type- vs. token-based NLP indices, particularly in analyses of longitudinal data.

As with any research, this dissertation project as a whole has a number of limitations. First, the sample size of the L2 learner corpus was small. While the repeated measures of the longitudinal data increased the power of statistical analysis, ideally a larger pool of L2 learners would have been recruited. Another important limitation was the fact that all experimental tasks were completed online at a time and location of the participants' choosing. While the accessibility of the tasks arguably allowed me to obtain more participants than would have been possible in a physical lab

In addition, while self-reported English proficiency was included as a control variable in all analyses, the initial proficiency of L2 learners was not homogenous. In order to obtain a desirable sample size, new IEP students were recruited who had recently placed into high beginner through high intermediate levels in their respective programs. It is certainly the case that variation in initial proficiency may have resulted in different rates of development across studies. Additionally, while L1 distance was included as a control variable across studies, the influence of learners' L1 was not properly accounted for and undoubtedly played a role in their processing and production of L2 English ((Ecke, 2015; Jarvis & Pavlenko, 2007).

7.1.4 Synthesis of Findings from Studies 1-3

Taken as a whole, studies 1-3 offer the following insight into the organization and development of L2 depth of knowledge from a lexicon-based perspective. First, there is no obvious single-best method for tapping into L2 lexical knowledge, a finding that has been acknowledged by others (e.g., Schmitt, 2010; Schmitt et al., 2001). In the current project, learners' WA behavior changed over time while their vocabulary size, evidence of semantic priming, and lexical production (as assessed by the dependent variables selected for analysis here) did not. These results demonstrate that the variety of ways in which lexical knowledge may be assessed each tell a different story about learner development and that parallel development should not be assumed either by researchers or assessment specialists. The fact that learners' vocabulary size failed to statistically predict other measures of lexical knowledge in Studies 1, 2, or 3 also suggest that breadth and depth of knowledge appear to be distinct constructs (Nurweni & Read, 1999)

These three studies also question the assumption that acquisition targets and benchmarks for learning should be derived from native-speakers' behavior at all (Klein, 1998; Ortega, 2013, 2016): Study 2 did not find evidence of learners' development of semantic priming as seen in NES data; Study 3 did not demonstrate change over time in measures of lexical production derived from native corpora; and Study 1 found that learners were becoming less native-like over time. While prominent SLA scholars like Ortega (2013, 2014, 2016) have recently begun to challenge (effectively monolingual) “native-likeness” as a target for SLA research, researchers outside of SLA (e.g., Canagarajah, 2006; Jenkins, 2002, 2005) have long problematized the use of native-speaker targets for L2 English learners given the role of English as an unofficial “lingua franca” (ELF) for international communication and the consequent implications for language policy, pedagogy, and assessment. Given the current project’s investigation of English specifically as an L2, the degree to which there exist multiple Englishes—many of them spoken by non-monolinguals—should be acknowledged (Canagarajah, 2006; Kachru, Kachru, & Nelson, 2009; McArthur, 2002; Seidlhofer, 2002, 2009). The fact that Studies 1-3 did not take learners’ goals for acquiring English into account (including the potential use of English as a lingua franca) constitutes a limitation of this dissertation project.

Finally, Studies 2 and 3 found no evidence of growth over time among measures of interest, supporting the notion that the development of depth of lexical L2 knowledge from a lexicon-based perspective is a slow and gradual process (Greidanus & Nienhuis, 2001; Henriksen, 1999, 2008). This is one reason why research has tended to neglect L2 speakers’ lexical networks, favoring investigations of form-meaning-mapping instead. It may that in order to observe the phenomena investigated in Studies 2 and 3, a longer longitudinal investigation would have been necessary. This seems especially likely in Study 2, where learners’ lexical decision behavior in response to association categories ran parallel to native-speakers’ but were not statistically significant.

In addition to those limitations acknowledged above, a major limitation of the methods used here is the assumption that the development of lexical knowledge can be observed independent from grammar and syntax. Such an assumption is increasingly challenged by usage-based approaches to learning (Bybee, 2008; N. C. Ellis, 2002b, 2008; Goldberg, 2002, 2006; Tomasello, 2003, 2007; Tyler, 2010) which maintain that the lexis and grammar are inter-related and do not constitute a unitary continuum (Gries & Stefanowitsch, 2007; Römer, 2009; Stemberger & MacWhinney, 1988).

7.2 Implications and Future Directions

Results from this project have some implications for our understanding of lexical acquisition and for further investigations of the L2 lexicon. First, as was demonstrated by L2 learners' word association's becoming less canonical over time (when compared to NES norms), the development that L2 learners do exhibit may not always be toward more "native-like" language use (Fitzpatrick, 2006). This finding suggests that using NES data as a default for baseline comparison in investigations of the L2 lexicon may also be problematic. Instead, researchers should consider comparing L2 learners to functionally bilingual and/or highly proficient L2 users, as has been advocated by SLA researchers such as Ortega (2013). SLA researchers should also be mindful of ELF scholarship (e.g., Canagarajah, 2006; Jenkins, 2002, 2005) which problematizes the use of English native-speaker targets in particular.

Findings across all three studies support the claim that breadth of lexical knowledge is distinct from depth of knowledge (e.g., Nurweni & Read, 1999), particularly from the manner in which depth of knowledge was operationalized and assessed in each of the three studies. Not only did vocabulary size fail to show growth over time in L2 learner models (see Tables 3.7 and 3..8), VLT scores were never retained as significant main effects in models predicting dependent measures overall (irrespective of time). Especially striking is the fact that VLT showed no impact on either L2 and UG (non-primed) lexical decision latencies (see Tables 5.9 and 5.12) More research, such as that conducted by Qian (1999, 2000) and Zareva (2005), is needed to unpack the relationship between breadth and depth of lexical knowledge and the extent to which the two may be interconnected or independent of one another at various stages of development. Given the fact that stimuli sublists factored into models even when those stimuli were balanced for lexical or psycholinguistic features, a methodological implication of this project is the importance of including demographic variables, individual differences, lexical features, and procedural variables in models.

In addition to the suggestions for future work outlined following individual research questions above, future research of the sort that comprised this dissertation project could include a mix of both quantitative and qualitative analysis across sub-studies to provide a more nuanced perspective of learners' development. For example, the exploratory post-hoc analysis in Study 1, which examined the later non-canonical word associations (WAs) of four L2 learners, might be enhanced by comparing qualitative analysis of learners' WAs to their lexical decision behavior or their lexical production. This manner of observing individual learners across studies might reveal divergent paths of non-discrete development that are other otherwise concealed by observing group behavior only (cf. Larsen-Freeman, 2006). Similarly, a mixed-method approach that combines quantitative and qualitative analysis (Hashemi & Babaii, 2013) could benefit projects like this one. For example, for data like those collected in Study 3, corpus tools could be used to identify formulaic or phraseological phenomena and qualitative analysis employed to determine the function of such language use.

7.3 Outlook

While more researchers have begun to employ innovative methods (including psycholinguistic approaches) to explore L2 lexical knowledge, few involve the analysis of multi-wave data collection over a significant amount of time (Ortega & Byrnes, 2008). This is especially true for L2 research that investigates lexical development and network building (Dóczi & Kormos, 2016; Henriksen, 1999). Because cross-sectional studies merely “describe a state of learners’ knowledge rather than a path of development” (Dóczi & Kormos, 2016, p. 29), more longitudinal studies like the ones reported here will allow researchers to capture change and examine the path of L2 lexical knowledge as it develops over time.

In particular, more empirical evidence is needed to gain insight into the manner in which theoretical conceptualizations of breadth and depth of L2 lexical knowledge develop with respect to one another within the same group of learners over time. With regard to network-based approaches to depth of knowledge specifically (Aitchison, 1994; Henriksen, 1999; Read, 2004), additional work will be necessary toward developing L2 measures that assess lexical items activated by their links to other items in the lexical network (Meara, 1997; Read, 2004).

For now, L2 lexical knowledge remains “many-faceted,” with no single best method capable of tapping into all of its forms (Schmitt et al., 2001a). It is hoped that the accumulation of evidence from a diversity of paradigms and methodological approaches will ultimately offer insights into understanding the multi-dimensional nature of L2 lexical knowledge and its development over time.

REFERENCES

Aitchison, J. (1994). *Language change: progress or decay?* New York: Springer Publishing.

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics*, *29*(2), 175-193.

Albert, M. S., Heller, H. S., & Milberg, W. (1988). Changes in naming ability with age. *Psychology and Aging*, *3*(2), 173-178.

Albrechtsen, D., Haastrup, K., & Henriksen, B. (2008). *Vocabulary and writing in a first and second language: Processes and development*. New York: Springer.

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419-439.

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 578-602.

Anderson, J. R. (1983a). The architecture of cognition. *Reprint, Mahwah, NJ: Lawrence Erlbaum Associates*.

Anderson, J. R. (1983b). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261-295.

Anderson, J. R. (1993). Rules of the Mind. LawrenceErlbaumAssociates: Hillsdale, New Jersey.

Anderson, J. R., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), *Comprehension and teaching: Research reviews.* (pp. 77-117). Newark, DE: International Reading Association.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database, release 2*.

Bagger Nissen, H., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics*, *16*(3), 389-408.

Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of verbal learning and verbal behavior, 22*(1), 88-104.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445-459.

Barcroft, J., Schmitt, N., & Sunderman, G. (2011). Lexis. *The Routledge handbook of applied linguistics*, 571-583.

Basnight-Brown, D. M., & Altarriba, J. (2007). Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory & Cognition, 35*(5), 953-965.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*.

Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing, 16*, 131-162.

Bell, H. (2009). The messy little details: A longitudinal case study of the emerging lexicon. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in second language learners. Papers* *and perspectives in honour of Paul Meara* (pp. 111-127). Bristol, UK: Multilingual Maters.

Bell, N., Skalicky, S., & Salsbury, T. (2014). Multicompetence in L2 language play: A longitudinal case study. *Language Learning*, *64*(1), 72-102.

Belz, J. A. (2002). Second language play as a representation of the multicompetent self in foreign language study. *Journal of Language, Identity, and Education*, *1*(1), 13-39.

Berger, C. M., Crossley, S. A., & Kyle, K. (2017a). Using native-speaker psycholinguistic norms to predict lexical proficiency and development in second-language production. *Applied Linguistics*.

Berger, C. M., Crossley, S. A., & Kyle, K. (2017b). Using novel word context measures to predict human ratings of lexical proficiency. *Journal of Educational Technology & Society*, *20*(2), 201-212.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*(2-3), 150-177.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition*, *16*(02), 157-168.

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. *Language and Computers*, *26*, 181-190.

Bleasdale, F. A. (1987). Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 582-594.

Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart & Winston.

Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition*, *23*(03), 321-343.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904-911.

Bybee, J. (2008). Usage-based grammar and second language acquisition. In *Handbook of cognitive linguistics and second language acquisition* (pp. 226-246). New York: Routledge.

Canagarajah, A. S. (2006). Negotiating the local in English as a lingua franca. *Annual Review of Applied Linguistics, 26,* 197-218.

Chapelle, C. A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research, 10*(2), 157-187.

Chapelle, C. A. (1998). JL Construct definition and validity inquiry in SLA research. *Interfaces between second language acquisition and language testing research*, 32.

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & psychophysics*, *36*(4), 387-395.

Chen, D., & Manning, C. (2014). A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 740-750).

Cheng, H. W. (2009). An overview of the research on lexical knowledge in China. *Sino-US English Teaching*, *6*(9), 52-57.

Chomsky, N. (1965). Aspects of the theory of syntax Cambridge. *Multilingual Matters: MIT Press*.

Chomsky, N. (1981). Principles and parameters in syntactic theory. *Explanation in linguistics: The logical problem of language acquisition, 32*, 75.

Clear, J. (1993). From Firth principles: Computational tools for the study of collocation. *Text and technology: In honour of John Sinclair*, 271-292.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407-428.

Cowie, A. P. (1998). *Phraseology: Theory, Analysis, and Applications: Theory, Analysis, and Applications*: Clarendon Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238.

Cronbach, L. J. (1942). An analysis of techniques for diagnostic vocabulary testing. *The journal of educational research*, *36*(3), 206-217.

Crossley, S. A. (2013). Assessing automatic processing of hypernymic relations in first language speakers and advanced second language learners: A semantic priming approach. *The Mental Lexicon*, *8*(1), 96-116.

Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*, *41*(4), 965-981.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35*(2), 115-135.

Crossley, S. A., & McNamara, D. (2013). Applications of text analysis tools for spoken response grading. *Language Learning & Technology*, *17*(2), 171-192.

Crossley, S. A., & Salsbury, T. L. (2011). The development of lexical bundle accuracy and production in English second language speakers. *IRAL-International Review of Applied Linguistics in Language Teaching*, *49*(1), 1-26.

Crossley, S. A., Salsbury, T., & McNamara, D. (2009). Measuring L2 lexical growth using hypernymic relationships. *Language Learning*, *59*(2), 307-334.

Crossley, S. A., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning, 60*(3), 573-605.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2011). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, *28*(4), 561-580.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2013). Validating lexical measures using human scores of lexical proficiency. In S. Jarvis & M. Daller (Eds.), *Vocabulary Knowledge: Human ratings and automated measures* (Vol. 47, pp. 105-135). Amsterdam: John Benjamins.

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2010). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*.

Crossley, S. A., Salsbury, T., Titak, A., & McNamara, D. S., (2014). Frequency effects and second language lexical acquisition: Word types, word tokens, and word production. *International Journal of Corpus Linguistics, 19* (3), 301-332.

Crossley, S. A., & Skalicky, S. (in press). Making sense of polysemy relations in first and second language speakers of English. *International Journal of Bilingualism.*

Crossley, S. A., Skalicky, S., Kyle, K., & Vanderbilt, K. (in press). Absolute frequency effects in second language lexical acquisition. *Studies in Second Language Acquisition*.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.

Csillagh, V. (2015). Global trends and local realities: Lessons about economic benefits, selves and identity from a Swiss context. *Studies in Second Language Learning and Teaching, 5*(3).

Csillagh, V. (2016). Love or money? Reinterpreting traditional motivational dimensions in modern social and economic contexts. In C. Gkonou, D. Tatzl & S. Mercer (Eds.), *New directions in language learning psychology* (pp. 185-208). Dördrecht: Springer.

Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*: Cambridge University Press Cambridge.

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied linguistics, 24*(2), 197-222.

Daller, M., Turlik, J., & Weir, I. (2013). Vocabulary acquisition and the learning curve. *Vocabulary Knowledge: Human ratings and automated measures, 47,* 185.

David, A. (2008). A developmental perspective on productive lexical knowledge in L2 oral interlanguage. *Journal of French Language Studies, 18*(03), 315-331.

Davies, Mark. (2009) *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Available online at https://corpus.byu.edu/coca/.

De Saussure, F. (1916). *Cours de linguistique générale*. Paris: Payot.

DeKeyser, R. M. (1996). Exploring automatization processes. *TESOL Quarterly, 30*(2), 349-357.

DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning an dpracticing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in second language acquisition* (pp. 42-63). Cambridge: Cambridge University Pres.

DeKeyser, R. M. (2007). Introduction: Situating the concept of practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1-18). New York: Cambridge University Press.

Devitto, Z., & Burgess, C. (2004). Theoretical and methodological implications of language experience and vocabulary skill: Priming of strongly and weakly associated words. *Brain and Cognition*, *55*(2), 295-299.

Dóczi, B., & Kormos, J. (2016). *Longitudinal developments in vocabulary knowledge and lexical organization*. New York: Oxford University Press.

Drum, P. A., & Konopak, B. (1987). Learning word meanings from written contexts. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition*. Hillsdale, NJ: Lawrence Erlbaum.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. New York: Pearson Assessments.

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, *15*(4), 850-855.

Ecke, P. (2015). Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingualism: Language and Cognition*, *18*(2), 145-162.

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 515.

Ellis, N. C. (1994). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 211-282). London: Academic Press.

Ellis, N. C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning*, *8*(2-3), 103-128.

Ellis, N. C. (2002a). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*(02), 143-188.

Ellis, N. C. (2002b). Reflections on frequency effects in language processing. *Second Language Acquisition, 24*, 297-339.

Ellis, N. C. (2005). At the inferface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*, 305-352.

Ellis, N. C. (2007). Language acquisition as rational cue-contingency learning. *Applied Linguistics* 27(1). 1-24.

Ellis, N. C. (2008). Usage-based and form-focused SLA: The implicit and explicit learning of constructions. *Language in the context of use: Cognitive approaches to language and language learning*, 93-1205.

Ellis, R. (1993). The structural syllabus and second language acquisition. *TESOL Quarterly, 27*(1), 91-113.

Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied linguistics, 16*(4), 409-441.

Ellis, R. (2004). The Definition and Measurement of L2 Explicit Knowledge. *Language Learning, 54*(2), 227-275. doi:10.1111/j.1467-9922.2004.00255.x

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language. *Studies in Second Language Acquisition, 27*(2), 141-172.

Ellis, R. (2009a). Implict and explicit learning, knowledge and instruction. In R. Ellis, S. Lowewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3-26). Bristol, UK: Multilingual Matters.

Ellis, R. (2009b). Retrospect and prospect. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 335-353). Bristol, UK: Multilingual Matters.

Ellis, R., Lowewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (Eds.). (2009). *Implicit and explicit knowledge in second language learning, testing and teaching* (Vol. 42). Bristol, UK: Multilingual Matters.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive science, 33*(4), 547-582.

Evert, S. (2008). Corpora and collocations. *Corpus linguistics. An international handbook, 2*, 1212-1248.

Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal, 31*(2), 105-119.

Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in the first-and second-language reading of fluent bilinguals. *Memory & Cognition, 11*(6), 565-574.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, CA: Brooks Cole.

Fitzpatrick, T. (2006). Habits and rabbits: Word associations and the L2 lexicon. *EUROSLA Yearbook, 6*, 121-145.

Fitzpatrick, T. (2012a). Tracking the changes: Vocabulary acquisition in the study abroad context. *Language Learning Journal, 41*, 81-98.

Fitzpatrick, T. (2012b). Word associations. *The Encyclopedia of Applied Lingusitics.*

Fitzpatrick, T., & Izura, C. (2011). Word association in L1 and L2: An exploratory study of response types, response times, and interlingual mediation. *Studies in Second Language Acquisition, 33*(03), 373-398.

Fitzpatrick, T., Playfoot, D., Wray, A., & Wright, M. J. (2013). Establishing the reliability of word association data for investigating individual and group differences. *Applied Linguistics*, *36*(1), 23-50.

Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory & Cognition*, *2*(2), 337-339.

Fowler, C. A., Wolford, G., Slade, R., & Tassinary, L. (1981). Lexical access with and without awareness. *Journal of Experimental Psychology: General*, *110*(3), 341-362.

Frenck, C., & Pynte, J. (1987). Semantic representation and surface forms: A look at across-language priming in bilinguals. *Journal of Psycholinguistic Research*, *16*(4), 383-396.

Frenck-Mestre, C., & Prince, P. (1997). Second language autonomy. *Journal of memory and language*, *37*(4), 481-501.

Gardner, R. C., Lalonde, R. M., Moorcroft, R., & Evers, F. T. (1987). Second language attrition: The role of motivation and use. *Journal of Language and Social Psychology, 6*, 29–47.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*(2), 256-281.

Goldberg, A. E. (2002). *Construction grammar*: Wiley Online Library.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*: Oxford University Press on Demand.

Grabois, H. (1999). The convergence of sociocultural theory and cognitive lingusitics. Lexical semantics and the L2 acquisition of love, fear and happiness. In G. B. Palmer & D. J. Occhi (Eds.), *Cultural constructions of emotional substrates*. Amsterdam: John Benjamins.

Granger, S., & Meunier, F. (2008). *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins Publishing.

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. *Phraseology. An interdisciplinary perspective*, 27-50.

Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, *18*(1), 109-136.

Greidanus, T., & Nienhuis, L. (2001). Testing the quality of word knowledge in a second language by means of word associations: Types of distractors and types of associations. *The Modern Language Journal*, *85*(4), 567-577.

Gries, S. T., & Stefanowitsch, A. (2007). *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis* (Vol. 1): Walter de Gruyter.

Harley, T. A. (2008). *The psychology of language: From data to theory* (3rd ed.). East Sussex, UK: Psychology Press.

Hazenberg, S., & Hulstijn, J. H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied linguistics*, *17*(2), 145-163.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, *21*(2), 303-317.

Henriksen, B. (2008). Declarative lexical knowl-edge. In D. Albrechtsen, K. Haastrup, and B. Henriksen (Eds.), *Vocabulary and writing in a first and secondlLanguage: Processes and development* (pp. 22-66). Lond: Palgrave Macmillan.

Hirsh, K. W., & Tree, J. J. (2001). Word association norms for two cohorts of British adults. *Journal of neurolinguistics*, *14*(1), 1-44.

Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718-730.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied linguistics*, *19*(1), 24-44.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23-B32.

Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica*, *121*(1), 65-80.

Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes-no vocabulary test: Correction for guessing and response style. *Language Testing*, *19*(3), 227-245.

Hulstijn, J. H. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, *18*(3), 193-223. doi:10.1191/0267658302sr207oa

Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, *8*(3), 229-249.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785-813.

Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C. S., ... & Buchanan, E. (2013). The semantic priming project. *Behavior Research Methods*, *45*(4), 1099-1114.

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(2), 130-136.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, *19*(1), 57-84.

Jarvis, S., & Pavlenko, A. (2007). *Cross-linguistic influence in language and cognition*. New York: Routledge.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an International Language. *Applied Linguistics*, *23*(1), 83-103.

Jenkins, J. (2005). Implementing an international approach to English pronunciation: The role of teacher attitudes and identity. *TESOL Quarterly, 39*(3), 535-43.

Jiang, N. (2012). *Conducting reaction time research in second language studies*. New York: Routledge.

Kachru, B., Kachru, Y., & Nelson, C. (eds.) (2009). *The Handbook of World Englishes* (vol. 48). Oxford: Wiley-Blackwell.

Kahan, T. A., Neely, J. H., & Forsythe, W. J. (1999). Dissociated backward priming effects in lexical decision and pronunciation tasks. *Psychonomic Bulletin & Review*, *6*(1), 105-110.

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic inquiry*, *8*(1), 63-99.

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitken, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies* (pp. 153-165). Edinburgh: Edinburgh University Press.

Klein, W. (1998). The contribution of second language acquisition research. *Language learning*, *48*(4), 527-549.

Kotz, S. A. (2001). Neurolinguistic evidence for bilingual language representation: A comparison of reaction times and event-related brain potentials. *Bilingualism: Language and Cognition*, *4*(2), 143-154.

Kotz, S. A., & Elston-Güttler, K. (2004). The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials. *Journal of Neurolinguistics*, *17*(2-3), 215-235.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Oxford: Pergamon.

Krashen, S. (1991). *The input hypothesis: An update.* Paper presented at the Georgetown University Round Table on Languages and Lingusitics.

Kruse, H., Pankhurst, J., & Smith, M. S. (1987). A multiple word association probe in second language acquisition research. *Studies in Second Language Acquisition*, *9*(02), 141-154.

Kucera, H. (1982). The mathematics of language. *The American heritage dictionary*, 37-41.

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication.* (Doctoral dissertation), Georgia State University. Retrieved from http://scholarworks.gsu.edu/alesl_diss/35

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*(4), 757-786.

Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods*, *50*(3), 1030-1046.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2-3), 259-284.

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 579-589.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics, 18*(2), 141-165.

Larsen-Freeman, D., & Cameron, L. (2008). Research methodology on language development from a complex systems perspective. *The Modern Language Journal*, *92*(2), 200-213.

Laufer, B. (1992). How much lexis is necessary for reading comprehension?. In *Vocabulary and applied linguistics* (pp. 126-132). Palgrave Macmillan, London.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied linguistics, 19*(2), 255-271.

Laufer, B. (2005a). Focus on form in second language vocabulary learning. *EUROSLA Yearbook, 5*(1), 223-250.

Laufer, B. (2005b). Lexical frequency profiles: From Monte Carlo to the real world. A response to Meara (2005). *Applied Linguistics, 26*(4), 582-588.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing*, *21*(2), 202-226.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399-436.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics*, *16*(3), 307-322.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33-51.

Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *EUROSLA Yearbook*, *1*(1), 7-28.

Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of languagelearning context. *Language Learning*, *48*(3), 365-391.

Lee, N. (2004). The neurobiology of procedural memory. In J. H. Schumann (Ed.), *The neurobiology of learning: Perspectives from second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25-42). Ann Arbor: University of Michigan Press.

Leow, R. P. (2015). Implicit learning in SLA: Of processes and products. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 47-66). Amsterdam: John Benjamins.

Leow, R. P., Grey, S., Marijuan, S., & Moorman, C. (2014). Concurrent data elicitation procedures, processes, and the early stages of L2 learning: A critical overview. *Second Language Research*, *30*(2), 111-127. doi:10.1177/0267658313511979

Leung, J. H., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition*, *33*(01), 33-55.

Leung, J. H., & Williams, J. N. (2012). Constraints on implicit learning of grammatical form-meaning connections. *Language Learning*, *62*(2), 634-662.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences, 22*(01), 1-38.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review, 7*(4), 618-630.

Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment.* Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT press.

Marsden, E., Mackey A., & Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS Repository of Instruments for Research into Second Languages* (pp. 1-21). New York: Routledge.

McArthur, T. (2002). *Oxford Guide to World Englishes*. Oxford: Oxford University Press.

McCarthy, P. M., & Jarvis, S. (2013). From intrinsic to extrinsic issues of lexical diversity assesment. In S. Jarvis & M. Daller (Eds.), *Vocabulary Knowledge: Human ratings and automated measures* (Vol. 45-78, pp. 45). Amsterdam: John Benjamins.

McDonough, K., & Trofimovich, P. (2009). *Using priming methods in second language research*. New York: Routledge.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York: Psychology Press.

Meara, P. (1978). Learners' word associations in French. *Interlanguage Studies Bulletin, 3*, 192-211.

Meara, P. (1980). *Vocabulary acquisition: A neglected aspect of language learning.* Paper presented at the Language teaching and linguistics: Abstracts.

Meara, P. (1983). Word associations in a foreign language. *Nottingham Linguistics Circular, 11*(2), 29-38.

Meara, P. (1984). The study of lexis in interlanguage. In A. Davies, C. Criper, & A. P. R. Howatt (Eds.), *Interlanguage* (pp. 225-235). Edinburgh: Edinburgh University Press.

Meara, P. (1992). *EFL vocabulary tests*. University College, Swansea: Centre for Applied Language Studies.

Meara, P. (1996). The dimensions of lexical competence. Performance and competence in second language acquisition: Cambridge: Cambridge University Press.

Meara, P. (1997). Towards a new aproach to modelling vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 109-121). Cambridge: Cambridge University Press.

Meara, P. (accessed 2008). *P_Lex v2.0: The Manual.* Available on the _*lognostics* website, http://www.lognostics.co.uk/.

Meara, P. (2009). *Connected words: Word associations and second language vocabulary association*. Amsterdam: John Benjamins.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4*(2), 142-154.

Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System, 28*(1), 19-30.

Meara, P., & Jones, G. (1990). Eurocentres vocabulary size test 10KA. *Zurich: Eurocentres*.

Meara, P. and Miralpeix, I. (accessed 2008). *V_Size: The Manual.* Available on the_lognostics website, http://www.lognostics.co.uk/.

Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 84-102). Cambridge: Cambridge University Press.

Meurers, D. (2013). Natural language processing and language learning. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-13): Blackwell Publishing Ltd.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.

Miller, E. N., & Chapman, L. J. (1983). Continued word association in hypothetically psychosis-prone college students. *Journal of Abnormal Psychology, 92*(4), 468-478.

Milton, J. (2009). *Measuring second language vocabulary acquistion*. Bristol, UK: Multilingual Matters.

Miller, G. A., & Fellbaum, C. (1991). Semantic networks of English. *Cognition, 41*(197-229).

Milton, J., & Fitzpatrick, T. (Eds.). (2014). *Dimensions of vocabulary knowledge*. Basingstoke, UK: Palgrave Macmillan.

Miralpeix, I. (2008). *The influence of age on vocabulary acquisition in English as a Foreign Language*. PhD thesis, Universitat de Barcelona.

Mochida, K., & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing, 23*(1), 73-98. doi:10.1191/0265532206lt321oa

Moon, R. (1997). *Vocabulary connections: Multiword items in English.* . Cambridge: Cambridge University Press.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 116.

Namei, S. (2004). Bilingual lexical development: A Persian-Swedish word association study. *International Journal of Applied Linguistics, 14*(3), 363-388.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines, 5*, 12-25.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle & Heinle Publishers.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. (2006). How large a vocabulary is needed for reading and listening?. *Canadian Modern Language Review*, *63*(1), 59-82.

Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques.* Boston, MA: Heinle.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9-13.

Nation, I. S. P., & Gu, P. Y. (2007). *Focus on vocabulary*. Sydney: National Centre for English Language Teaching and Research.

Nation, P., & Heatley, A. (1994). *VocabProfile: A program for analyzing vocabulary in texts.* Wellington, New Zealand: Victoria University of Wellington.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition*, *4*(5), 648-654.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226.

Neely, J. H. (2012). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In *Basic processes in reading* (pp. 272-344). New York: Routledge.

Neely, J. H. (2012). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In *Basic processes in reading* (pp. 272-344). Routledge.

Nelson, D., McEvoy, C., & Dennis, S. (2000). What is and what does free association measure. *Memory & Cognition*, *28*(6), 887-899.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402-407.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition, 1*, 1-55.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555-578.

Nurweni, A., & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes, 18*(2), 161-175.

O'Grady, W. (2006). *The problem of verbal inflection in second language acquistion*. Paper presented at the Paper presented at the 11th PAAL Conference, Chuncheon, Korea.

Orita, M. (2002). Word associations of Japanese EFL learners and native speakers: Shifts in response type distribution and the associative development of individual words. *Annual Review of English Language Education in Japan, 13*, 111-120.

Ortega, L. (2016). Multi-competence in second language acquisition: inroads into the mainstream?. In V. Cook & L. Wei (Eds,), *The Cambridge Handbook of Linguistic Multi competence* (pp. 50-76). Cambridge University Press.

Ortega, L. (2013). SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning*, *63*, 1-24.

Ortega, L., & Byrnes, H. (2008). The longitudinal study of advanced L2 capacities: An introduction. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 3-20). New York: Routledge.

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics, 25*, 26-45.

Paciorek, A., & Williams, J. N. (2015). Semantic implicit learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 67-89). Amsterdam: John Benjamins.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*(3), 255.

Paradis, M. (2000). Cerebral representation of bilingual concepts. *Bilingualism: Language and Cognition*, *3*(01), 22-24.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins Publishing.

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. *Second language vocabulary acquisition: A rationale for pedagogy*, *55*(4), 174-200.

Pecina, Pavel. 2009. Lexical association measures and collocation extraction. *Language Resources and Evaluation* 44(1-2). 137-158.

Phillips, N. A., Segalowitz, N., O'Brien, I., & Yamasaki, N. (2004). Semantic priming in a first and second language: Evidence from reaction time variability and event-related brain potentials. *Journal of Neurolinguistics*, *17*(2-3), 237-262.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied linguistics*, *10*(1), 52-79.

Pinker, S. (1991). Rules of language. *Science*, *253*(5019), 530-535.

Pollock, J.-Y. (1989). Verb movement, Universal Grammar, and the structure of IP. *Linguistic inquiry*, *20*(3), 365-424.

Postman, L. J., & Keppel, G. (1970). *Norms of Word Association*. Academic Press.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian modern language review*, *56*(2), 282-308.

Qian, D. D. (2000). *Validating the Role of Depth of Vocabulary Knowledge in Assessing Reading for Basic Comprehension*. Retrieved from Princeton, NJ:

Randall, M. (1980). Word association behavior in learners of English as a second language. *Polyglot*, *2*(2), B4-D1.

Randall, M. (2007). *Memory, psychology and second language learning* (Vol. 19): John Benjamins Publishing.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology Section A*, *55*(4), 1339-1362.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*(1), 159–182.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*(3), 355-371.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment*. Mahwah, NJ: Lawrence Erlbaum.

Read, J. (2000). *Assessing vocabulary*: Cambridge University Press.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a Second Language* (pp. 209-228). Amsterdam: John Benjamins Publishing.

Read, J. (2005). Applying lexical statistics to the IELTS speaking test. *Research Notes, 20*, 12-16.

Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing, 18*(1), 1-32.

Ready-Morfitt, D. (1991). *The role and limitations of self-assessment in testing and research. Unpublished manuscript*. University of Ottawa: Ottawa, Canada.

Revier, R. L. (2009). Evaluating a new test of *whole* English collocations. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 125-138). Basingstoke, UK: Palgrave Macmillan.

Riegel, K. F., & Zivian, I. W. (1972). A study of inter-and intralingual associations in English and German. *Language Learning*, *22*(1), 51-63.

Römer, U. (2009). The inseparability of lexis and grammar Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics, 7*(1), 140-162.

Römer, U., O'Donnell, M. B., & Ellis, N. C. (2014). Second language learner knowledge of verb–argument constructions: Effects of language transfer and typology. *The Modern Language Journal, 98*(4), 952-975.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Memory and Language*, *9*(5), 487-497.

Sanz, C., & Grey, S. (2015). Effects of condition on L2 development: Moving beyond accruacy. In P. Rebuschat (Ed.), *Implict and explicit learning of languages* (pp. 301-324). Amsterdam: John Benjamins.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human perception and performance*, *3*(1), 1-17.

Scarcella, R., & Zimmerman, C. (1998). Academic words and gender. *Studies in Second Language Acquisition*, *20*(01), 27-49.

Schiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing, II: Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan.

Schmitt, N. (1998). Quantifying word association responses: What is native-like?. *System*, *26*(3), 389-401.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action. *Formulaic sequences*, 1-22.

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge network: Word associations and verbal sufixes. *Studies in Second Language Acquisition*, *19*(01), 17-36.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55-88.

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, 145-171.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*(5), 499-520.

Schwanenflugel, P. J., & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. *Journal of Memory and Language*, *25*(5), 605-618.

Sedivy, J. S. (2010). Using eyetracking in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 115-138). Amsterdam: John Benjamins.

Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, *19*(01), 53-67.

Seidlhofer, B. (2002). Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics, 11*(2), 133-58.

Selinker, L., & Gass, S. (2008). *Second language acquisition: An introductory course*: Taylor & Francis.

Service, E. T. (1987). TOEFL test and score manual. Princeton, NJ: Author.

Singleton, D. M. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.

Singleton, D. M. (2000*). Language and the lexicon*. London: Arnold.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian modern language review, 64*(3), 429-458.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510-532.

Söderman, T. (1993). Word association results: A window to the lexicons of ESL students. *JALT Journal, 15*(2), 135-150.

Sökmen, A. (1993). Word associaiton results: A window to the lexicons of ESL students. *JALT Journal*, *15*(2), 135-150.

Stemberger, J. P., & MacWhinney, B. (1988). *Are inflected forms stored in the lexicon?* In M. Hammond & M. Noonan (Eds.), *Theoretical morphology: Approaches in modern linguistics* (pp. 101–116). San Diego, CA: Academic Press.

Tomasello, M. (2003). Constructing a language: A usage-based approach to child language acquisition: Cambridge, MA: Harvard University Press.

Tomasello, M. (2007). Cognitive Linguistics and first language acquisition. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford handbook of cognitive linguistics* (pp. 1092-1112). Oxford: Oxford University Press.

Tyler, A. (2010). Usage-based approaches to language and their applications to second language learning. *Annual Review of Applied Linguistics, 30*, 270-291.

van Ginkel, C. I., & van der Linden, E. H. (1996). Word associations in foreign language learning and foreign language loss. In K. Sajavaara & C. Fairweather (Eds.), *Approaches to second language acquisition*. Jyväskylä: University of Jyväskylä.

Vasos, H. (1983). *Semantic processing in bilinguals.* Unpublished PhD dissertation. Concordia University. Montreal, Quebec.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics, 22*(02), 217-234.

Verspoor, M., De Bot, K., & Lowie, W. (Eds.). (2011). *A dynamic approach to second language development: Methods and techniques* (Vol. 29). Amsterdam: John Benjamins Publishing.

Wesche, M., & Paribakht, T. S. (1996). Assessing Second Language Vocabulary Knowledge: Depth Versus Breadth. *Canadian Modern Language Review, 53*(1), 13-40.

Wesche, M. B., & Paribakht, T. (2009). *Lexical inferencing in a first and second language: Cross-linguistic dimensions* (Vol. 46): Multilingual Matters.

West, M. P. (Ed.). (1953). *A general service list of English words: With semantic frequencies and a supplementary word-list for the writing of popular science and technology*. Longmans, Green: Addison-Wesley Longman Limited.

Wolter, B. (2002). Assessing proficiency through word associations: Is there still hope?, *System, 30,* 315–29.

Wray, A. (2002). Formulaic language and the lexicon: Cambridge: Cambridge University Press.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1.

Zareva, A. (2005). Models of lexical knowledge assessment of second language learners of English at higher levels of language proficiency. *System, 33*(4), 547-562.

Zareva, A., & Wolter, B. (2012). The "promise" of three methods of word association analysis to L2 lexical research, *Second Language Research, 28*(1), 41–67.

Zhang, X., & Lu, X. (2013). A longitudinal study of receptive vocabulary breadth knowledge growth and vocabulary fluency development. *Applied linguistics*.

Zipf, G. K. (1935). *The psycho-biology of language*. Boston, MA: Houghton Mifflin.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Reading, UK: Addison-Wesely.

APPENDICES

Appendix A: Demographic Questionnaire

1. What is your gender?
   - o Male
   - o Female
   - o Other ____________________

2. Which hand do you use to write with?
   - o Left
   - o Right
   - o I use both hands to write

3. What is your age? Please type your age as a whole number (for example, 22):

4. Which of the following applies to you?
   - o I am a student in an Intensive English Program (IEP) (1)
   - o I am an undergraduate student
   - o I am a graduate or post-baccalaureate student
   - o Other: ____________________

5. What is your native language(s)?

This is commonly the language (or languages) that you learned in your home as a child, though some bilingual speakers may have learned a dominant language in school at a very young age. You may list more than one native language. You may also use this space to explain your relationship to a language (for example, “English: My family did not speak English, but I learned it in preschool and consider myself bilingual.”)

** Participants only encountered Questions 6-8 if they indicated that they were IEP students.*

6. How long have you lived in the United States or another English-speaking country?
   - o I have lived in an English-speaking country my whole life
   - o One month or less
   - o Three months or less
   - o Six months or less
   - o One year or less
   - o Two years or less
   - o Five years or less
   - o More than five years
   - o Other ____________________

7. List the year of your first arrival to an English-speaking country that lasted three months or longer (for example, “2014”):

If you have never been to an English-speaking country for more than 1-2 months, write “N/A”.

8. Please list the month and year you began your IEP program in the United States (for example, “January 2017”):

** Participants only encountered Questions 9-11, if they indicated a language(s) other than English in Question 5.*

9. Is there anything else about your relationship to English that you would like to share? It is OK to skip this question.

10. Location of secondary education:
Please respond with a country name. If you went to high school in the United States, you should simply write "United States."

11. Language(s) of secondary education:
Please write the language that your instructors used to teach content courses. If you went to school in the United States, this was probably English (unless you went to a bilingual or international high school).

12. What is your major or area of study?
If this question does not apply, you can skip this question.

13. Please list any additional languages you know (other than English or your native language). If you do not know any other languages, you can leave this section blank.

- Language __________________
- Language __________________
- Language __________________
- Language __________________
- Language __________________
- Language __________________

14. In {Language X}, I can speak and understand...

- simple, everyday expression and phrases
- clear descriptions and formulated language that function in most situations
- spontaneous and complex conversations and readings
- practically any form of the language with nuance and precision

** Question 14 iterated based on the language(s) listed in Question 23, if any.*

Appendix B: Language Experience Survey

1. In English, I can speak and understand…
   - o simple, everyday expressions and phrases
   - o clear descriptions and formulated language that function in most situations
   - o spontaneous and complex conversations and readings
   - o practically any form of the language with nuance and precision

2. I use English with friends and acquaintances in the United States…
   - o Never or almost never
   - o Sometimes (about once a month)
   - o Often (once a week)
   - o Very often (every day)

3. I use English at work in the United States...
   - o I do not work
   - o Never or almost never
   - o Sometimes (about once a month)
   - o Often (once a week)
   - o Very often (every day)

4. I use English... (check all that apply)
   - o To read
   - o To write
   - o For listening
   - o For speaking

5. I use English to write emails...
   - o I do not write emails
   - o Never or almost never
   - o Sometimes (about once a month)
   - o Often (once a week)
   - o Very often (every day)

6. I use English to participate in social media (Facebook, Twitter, SnapChat, Instagram, etc.)...
   - o I do not participate in social media
   - o Never or almost never
   - o Sometimes (about once a month)
   - o Often (once a week)
   - o Very often (every day)

7. I use English on the Internet...
   - o Never or almost never
   - o Sometimes (about once a month)
   - o Often (once a week)
   - o Very often (every day)

8. I use English to write/read text messages...
    - I do not send text messages
    - Never or almost never
    - Sometimes (about once a month)
    - Often (once a week)
    - Very often (every day)

9. I use English to watch films or television…
    - I do not watch films or television
    - Never or almost never
    - Sometimes (about once a month)
    - Often (once a week)
    - Very often (every day)

10. I read books in English...
    - I do not read books
    - Never or almost never
    - Sometimes (about once a month)
    - Often (once a week)
    - Very often (every day)

11. I read magazines or newspapers in English...
    - I do not read magazines or newspapers
    - Never or almost never
    - Sometimes (about once a month)
    - Often (once a week)
    - Very often (every day)

12. I use English to read the news on Internet websites...
    - I do not read the news on the Internet
    - Never or almost never
    - Sometimes (about once a month)
    - Often (once a week)
    - Very often (every day)

13. How long have you lived in the United States or another English-speaking country?
    - I have lived in an English-speaking country my whole life
    - One month or less
    - Three months or less
    - Six months or less
    - One year or less
    - Two years or less
    - Five years or less
    - More than five years
    - Other ______________________

14. Which situation best describes your current living situation?
   - I live in a student residence with other students
   - I live in a student resident with other IEP or international students
   - I live alone in an apartment or house
   - I live in an apartment or house with my family
   - I live in an apartment or house with people from the United States
   - I live in an apartment or house with other international people
   - Other ____________________

15. On average, how many days per week do you spend speaking English? (Do not count English classes.)
   - 0
   - 1
   - 2
   - 3
   - 4
   - 5
   - 6
   - 7

16. On those days, how many hours per day do you usually spend speaking English?
   - 0-1
   - 1-2
   - 2-3
   - 3-4
   - 4-5
   - more than 5

17. My friends in the United State are…
   - mostly other IEP or international students
   - mostly other Americans
   - a mix of Americans and international students
   - Other ____________________

18. Is there anything else about your relationship to English that you would like to share? It is OK to skip this question.

Appendix C: Language Experience Survey Scoring

The following three measures were calculated as independent variables for models in Chapters 4-6: *self-reported proficiency*, *use*, and *exposure*. How each was calculated based on Language Experience Survey results is indicated below.

Self-reported proficiency

In English, I can speak and understand…	
Subject's response	Score reported in variable
simple, everyday expressions and phrases	1
clear descriptions and formulated language that function in most situations	2
spontaneous and complex conversations and readings	3
practically any form of the language with nuance and precision	4

Use: Number of hours using English per day

How many hours per day do you usually spend speaking English?	
Subject's response	Score reported in variable
0-1	1
1-2	2
2-3	3
3-4	4
4-5	5
more than 5	6

Exposure to English

For *exposure*, the average of subjects' responses to the 10 questions below was used to calculate a single score analyzed at each data collection.

Subject's response	Score used to calculate mean
I use English with friends and acquaintances in the U.S.	
Never or almost never	0
Sometimes (about once a month)	1
Often (once a week)	2
Very often (every day)	3
N/A ("I do not ___")	0
I use English at work in the United States...	
Never or almost never	0
Sometimes (about once a month)	1

Often (once a week)	2
Very often (every day)	3
N/A ("I do not ___")	0
I use English to write emails...	
Never or almost never	0
Sometimes (about once a month)	1
Often (once a week)	2
Very often (every day)	3
N/A ("I do not ___")	0
I use English to participate in social media (Facebook, Twitter, SnapChat, Instagram, etc.)...	
Never or almost never	0
Sometimes (about once a month)	1
Often (once a week)	2
Very often (every day)	3
N/A ("I do not ___")	0
I use English on the Internet...	
Never or almost never	0
Sometimes (about once a month)	1
Often (once a week)	2
Very often (every day)	3
N/A ("I do not ___")	0
I use English to write/read text messages...	
Never or almost never	0
Sometimes (about once a month)	1
Often (once a week)	2
Very often (every day)	3
N/A ("I do not ___")	0
I use English to watch films or television…	
Never or almost never	0
Sometimes (about once a month)	1
Often (once a week)	2
Very often (every day)	3
N/A ("I do not ___")	0
I read books in English...	

Never or almost never	0
Sometimes (about once a month)	1
Often (once a week)	2
Very often (every day)	3
N/A ("I do not ___")	0
I read magazines or newspapers in English...	
Never or almost never	0
Sometimes (about once a month)	1
Often (once a week)	2
Very often (every day)	3
N/A ("I do not ___")	0
I use English to read the news on Internet websites...	
Never or almost never	0
Sometimes (about once a month)	1
Often (once a week)	2
Very often (every day)	3
N/A ("I do not ___")	0

Appendix D: Vocabulary Size Test (VLT) Sub-Versions

Sub-Version 1

2,000-word level

1 birth
2 dust ______ game
3 operation ______ winning
4 row ______ being born
5 sport
6 victory

1 adopt
2 climb ______ go up
3 examine ______ look at closely
4 pour ______ be on every side
5 satisfy
6 surround

1 original
2 private ______ first
3 royal ______ not public
4 slow ______ all added together
5 sorry
6 total

1 accident
2 debt ______ loud deep sound
3 fortune ______ something you must pay
4 pride ______ having a high opinion of yourself
5 roar
6 thread

1 arrange
2 develop ______ grow
3 lean ______ put in order
4 owe ______ like more than something else
5 prefer
6 seize

3,000-word level

1 belt
2 climate ______ idea
3 executive ______ inner surface of your hand

4 notion ______ strip of leather worn around the waist
5 palm
6 victim

1 betray
2 dispose ______ frighten
3 embrace ______ say publicly
4 injure ______ hurt seriously
5 proclaim
6 scare

1 apartment
2 candle ______ a place to live
3 draft ______ chance of something happening
4 horror ______ first rough form of something written
5 prospect
6 timber

1 annual
2 concealed ______ wild
3 definite ______ clear and certain
4 mental ______ happening once a year
5 previous
6 savage

1 administration
2 angel ______ group of animals
3 frost ______ spirit who serves a god
4 herd ______ managing business and affairs
5 fort
6 pond

5,000 word level

1 balloon
2 federation ______ bucket
3 novelty ______ unusual interesting thing
4 pail ______ rubber bag that is filled with air
5 veteran
6 ward

1 blend
2 devise ______ mix together
3 hug ______ plan or invent
4 lease ______ hold tightly in your arms
5 plague

6 reject

1 cavalry
2 eve ______ small hill
3 ham ______ day or night before a holiday
4 mound ______ soldiers who fight from horses
5 steak
6 switch

1 correspond
2 embroider ______ exchange letters
3 lurk ______ hide and wait for someone
4 penetrate ______ feel angry about something
5 prescribe
6 resent

1 decent
2 frail ______ weak
3 harsh ______ concerning a city
4 incredible ______ difficult to believe
5 municipal
6 specific

10,000 word level

1 auspices
2 dregs ______ confused mixture
3 hostage ______ natural liquid present in the mouth
4 jumble ______ worst and most useless parts of anything
5 saliva
6 truce

1 apparition
2 botany ______ ghost
3 expulsion ______ study of plants
4 insolence ______ small pool of water
5 leash
6 puddle

1 alabaster
2 chandelier ______ small barrel
3 dogma ______ soft white stone
4 keg ______ tool for shaping wood
5 rasp
6 tentacle

1 dissipate
2 flaunt ______ steal
3 impede ______ scatter or vanish
4 loot ______ twist the body about uncomfortably
5 squirm
6 vie

1 illicit
2 lewd ______ immense
3 mammoth ______ against the law
4 slick ______ wanting revenge
5 temporal
6 vindictive

Sub-Version 2

2,000 word level

1 choice
2 crop ______ heat
3 flesh ______ meat
4 salary ______ money paid regularly for doing a job
5 secret
6 temperature

1 cream
2 factory ______ part of milk
3 nail ______ a lot of money
4 pupil ______ person who is studying
5 sacrifice
6 wealth

1 bake
2 connect ______ join together
3 inquire ______ walk without purpose
4 limit ______ keep within a certain size
5 recognize
6 wander

1 brave
2 electric ______ commonly done
3 firm ______ wanting food
4 hungry ______ having no fear
5 loca
6 usual

1 clerk
2 frame ______ a drink
3 noise ______ office worker
4 respect ______ unwanted sound
5 theater
6 wine

3,000 word level

1 acid
2 bishop ______ cold feeling
3 chill ______ farm animal
4 ox ______ organization or framework
5 ridge
6 structure

1 encounter
2 illustrate ______ meet
3 inspire ______ beg for help
4 plead ______ close completely
5 seal
6 shift

1 dim
2 junior ______ strange
3 magnificent ______ wonderful
4 maternal ______ not clearly lit
5 odd
6 weary

1 assemble
2 attach ______ look closely
3 peer ______ stop doing something
4 quit ______ cry out loudly in fear
5 scream
6 toss

1 blanket
2 contest ______ holiday
3 generation ______ good quality
4 merit ______ wool covering used on beds
5 plot
6 vacation

5,000 word level

1 alcohol
2 apron _____ stage of development
3 hip _____ state of untidiness or dirtiness
4 lure _____ cloth worn in front to protect your clothes
5 mess
6 phase

1 casual
2 desolate _____ sweet-smelling
3 fragrant _____ only one of its kind
4 radical ____ good for your health
5 unique
6 wholesome

1 concrete
2 era _____ circular shape
3 fiber _____ top of a mountain
4 loop _____ a long period of time
5 plank
6 summit

1 artillery
2 creed _____ a kind of tree
3 hydrogen _____ system of belief
4 maple _____ large gun on wheels
5 pork
6 streak

1 contemplate
2 extract _____ think about deeply
3 gamble _____ bring back to health
4 launch _____ make someone angry
5 provoke
6 revive

10,000 word level

1 antics
2 batch _____ foolish behavior
3 connoisseur _____ a group of things
4 foreboding _____ person with a good knowledge of art or music
5 haunch
6 scaffold

1 clinch
2 jot _____ move very fast

3 mutilate _____ injure or damage
4 smolder _____ burn slowly without flame
5 topple
6 whiz

1 alcove
2 impetus _____ priest
3 maggot _____ release from prison early
4 parole _____ medicine to put on wounds
5 salve
6 vicar

1 contaminate
2 cringe _____ write carelessly
3 immerse _____ move back because of fear
4 peek _____ put something under water
5 relay
6 scrawl

1 indolent
2 nocturnal _____ lazy
3 obsolete _____ no longer used
4 torrid _____ clever and tricky
5 translucent
6 wily

Sub-Version 3

2,000 word level

1 cap
2 education _____ teaching and learning
3 journey _____ numbers to measure with
4 parent _____ going to a far place
5 scale
6 trick

1 burst
2 concern _____ break open
3 deliver _____ make better
4 fold _____ take something to someone
5 improve
6 urge

1 coffee
2 disease _____ money for work

3 justice ______ a piece of clothing
4 skirt ______ using the law in the right way
5 stage
6 wage

1 blame
2 elect ______ make
3 jump ______ choose by voting
4 manufacture ______ become like water
5 melt
6 threaten

1 ancient
2 curious ______ not easy
3 difficult ______ very old
4 entire ______ related to god
5 holy
6 social

3,000 word level

1 bench
2 charity ______ long seat
3 jar ______ help to the poor
4 mate ______ part of a country
5 mirror
6 province

1 assist
2 bother ______ help
3 condemn ______ cut neatly
4 erect ______ spin around quickly
5 trim
6 whirl

1 bull
2 champion ______ formal and serious manner
3 dignity ______ winner of a sporting event
4 hell ______ building where valuable objects are shown
5 museum
6 solution

1 atmosphere
2 counsel ______ advice
3 factor ______ a place covered with grass
4 hen ______ female chicken

5 lawn
6 muscle

1 brilliant
2 distinct ______ thin
3 magic ______ steady
4 naked ______ without clothes
5 slender
6 stable

5,000 word level

1 apparatus
2 compliment ______ expression of admiration
3 ledge ______ set of instruments or machinery
4 revenue ______ money received by the government
5 scrap
6 tile

1 abolish
2 drip ______ bring to an end by law
3 insert ______ guess about the future
4 predict ______ calm or comfort someone
5 soothe
6 thrive

1 bleed
2 collapse ______ come before
3 precede ______ fall down suddenly
4 reject ______ move with quick steps and jumps
5 skip
6 tease

1 circus
2 jungle ______ musical instrument
3 nomination ______ seat without a back or arms
4 sermon ______ speech given by a priest in a church
5 stool
6 trumpet

1 adequate
2 internal ______ enough
3 mature ______ fully grown
4 profound ______ alone away from other things
5 solitary
6 tragic

10,000 word level

1 casualty
2 flurry ______ someone killed or injured
3 froth ______ being away from other people
4 revelry ______ noisy and happy celebration
5 rut
6 seclusion

1 arsenal
2 barracks ______ happiness
3 deacon ______ difficult situation
4 felicity ______ minister in a church
5 predicament
6 spore

1 auxiliary
2 candid ______ bad-tempered
3 luscious ______ full of self-importance
4 morose ______ helping, adding support
5 pallid
6 pompous

1 benevolence
2 convoy ______ kindness
3 lien ______ set of musical notes
4 octave ______ speed control for an engine
5 stint
6 throttle

1 blurt
2 dabble ______ walk in a proud way
3 dent ______ kill by squeezing someone's throat
4 pacify ______ say suddenly without thinking
5 strangle
6 swagger

Sub-Version 4

2,000 word level

1 attack
2 charm ______ gold and silver
3 lack ______ pleasing quality
4 pen ______ not having something

5 shadow
6 treasure

1 copy	
2 event	______ end or highest point
3 motor	______ this moves a car
4 pity	______ thing made to be like another
5 profit	
6 tip	

1 dozen	
2 empire	______ chance
3 gift	______ twelve
4 opportunity	______ money paid to the government
5 relief	
6 tax	

1 bitter	
2 independent	______ beautiful
3 lovely	______ small
4 merry	______ liked by many people
5 popular	
6 slight	

1 admire	
2 complain	______ make wider or longer
3 fix	______ bring in for the first time
4 hire	______ have a high opinion of someone
5 introduce	
6 stretch	

3,000 word level

1 boot	
2 device	______ army officer
3 lieutenant	______ a kind of stone
4 marble	______ tube through which blood flows
5 phrase	
6 vein	

1 comment	
2 gown	______ long formal dress
3 import	______ goods from a foreign country
4 nerve	______ part of the body which carries feeling
5 pasture	
6 tradition	

1 abandon
2 dwell ______ live in a place
3 oblige ______ follow in order to catch
4 pursue ______ leave something permanently
5 quote
6 resolve

1 drift
2 endure ______ suffer patiently
3 grasp ______ join wool threads together
4 knit ______ hold firmly with your hands
5 register
6 tumble

1 aware
2 blank ______ usual
3 desperate ______ best or most important
4 normal ______ knowing what is happening
5 striking
6 supreme

5,000 word level

1 bulb
2 document ______ female horse
3 legion ______ large group of soldiers or people
4 mare ______ a paper that provides information
5 pulse
6 tub

1 gloomy
2 gross ______ empty
3 infinite ______ dark or sad
4 limp ______ without end
5 slim
6 vacant

1 analysis
2 curb ______ eagerness
3 gravel ______ loan to buy a house
4 mortgage ______ small stones mixed with sand
5 scar
6 zeal

1 chart

2 forge ______ map
3 mansion ______ large beautiful house
4 outfit ______ place where metals are made and shaped
5 sample
6 volunteer

1 demonstrate
2 embarrass ______ have a rest
3 heave ______ break suddenly into small pieces
4 obscure ______ make someone feel shy or nervous
5 relax
6 shatter

10,000 word level

1 blaspheme
2 endorse ______ slip or slide
3 nurture ______ give care and food to
4 skid ______ speak badly about a god
5 squint
6 straggle

1 dubious
2 impudent ______ rude
3 languid ______ very ancient
4 motley ______ of many different kinds
5 opaque
6 primeval

1 acquiesce
2 bask ______ to accept without protest
3 crease ______ sit or lie enjoying warmth
4 demolish ______ make a fold on cloth or paper
5 overhaul
6 rape

1 bourgeois
2 brocade ______ middle class people
3 consonant ______ row or level of something
4 prelude ______ cloth with a pattern or gold or silver threads
5 stupor
6 tier

1 alkali
2 banter ______ light joking talk

3 coop	_____ a rank of British nobility
4 mosaic	_____ picture made of small pieces of glass or stone
5 stealth	
6 viscous	

Appendix E: Framework for Categorizing Word Association Responses

Category	**Subcategory**	**Definition**	**Example**
Meaning-based association	Synonym	Cue and response are synonymous	*delay => impede*
	Antonym	Cue and response are antonyms	*permit => deny*
	Hierarchical set	Cue and response share a hyponym	*bean => vegetable* *vehicle => car*
	Coordination	Cue and response belong to the same category	*trumpet => piano*
	Conceptual other (quality association, contextual association, other)	One word is a quality of the other, one word gives a conceptual context for another, or some other conceptual link exists.	*elephant => big* *forest => bird* *sin => prayer* *party => celebrate*
Position-based association	Cue-response collocation	Cue is followed by the response in common usage; includes phrasal as well as consecutive collocation	*fence => post* *swear => word*
	Response-cue collocation	Response is followed by the word in common usage; includes phrasal as well as consecutive collocation	*fence => electric* *plug => spark*
	Bi-directional collocation	Cue can precede or follow response in common usage	*rock => hard*
Form-based association	Affix manipulation	Response adds, deletes, or changes an inflectional or derivational affix to cue	*irony => ironic* *plug => unplug*
	Formal features	Response looks like and/or sounds like cue but has no meaning relation (i.e., phonological or orthographic features of the cue word trigger the response)	*fence => hence* *wait => weight*
	Two-step form association	Cue and response are linked through form-based association with another word	*weak => Monday* (via *week*) *owe => mine* (via *own*)
Hodgepodge	Empty	No response is given but cue word is indicated as known	
	Repetition	Cue word is repeated	*beautiful => beautiful*
	Translation	Response is a translation of the cue word or response is given in another language	*table => mesa*
	Erratic	Link between cue and response seems illogical or is otherwise impossible to code	*lion => and*
Dual-coding	Form and meaning	Response and cue related in both form and general meaning	*hairdress => hairdryer*

	Meaning and position	Response and cue related in both general meaning and tendency to co-occur in language usage (includes uni- and bi-directional collocation)	*pearl* => *necklace* *brother* => *sister* *fork* => *spoon* *shove* => *push*

Note. Taxonomy adapted from those previously created by Albrechtsen, Haastrup, and Henriksen (2008); Fitzpatrick (2006); Fitzpatrick and Izura (2011); and Fitzpatrick, Playfoot, Wray, and Wright (2013)

Appendix F: Complete Stimuli for Study 2

								Prime		Target		
L2 subset	**NES subset**	**Target condition**	**Relation condition**	**Micro-relation**	**Prime**	**Target**	**FSG**	**Freq**	**Length**	**Freq**	**Length**	**Part of Speech**
L2_A	NS_1	real	meaning	antonym	*late*	*early*	0.47	86421	4	108171	5	aj
L2_A	NS_1	real	meaning	antonym	*exit*	*enter*	0.38	7333	4	54479	5	v
L2_A	NS_1	real	meaning	conceptual other	*bee*	*honey*	0.235	6342	3	11009	5	n
L2_A	NS_1	real	meaning	conceptual other	*calendar*	*date*	0.3	5258	8	31467	4	n
L2_A	NS_1	real	meaning	coordination	*square*	*circle*	0.47	11630	6	24735	6	n
L2_A	NS_1	real	meaning	coordination	*spoon*	*fork*	0.614	6194	5	7742	4	n
L2_A	NS_1	real	meaning	hierarchical set	*color*	*red*	0.242	56978	5	66217	3	n
L2_A	NS_1	real	meaning	hierarchical set	*robot*	*machine*	0.22	6023	5	38407	7	n
L2_A	NS_1	real	meaning	synonym	*dumb*	*stupid*	0.52	5486	4	12467	6	aj
L2_A	NS_1	real	meaning	synonym	*trim*	*cut*	0.497	5113	4	96012	3	v
L2_A	NS_1	real	position	cue-response	*tissue*	*paper*	0.28	12342	6	75383	5	n
L2_A	NS_1	real	position	cue-response	*shopping*	*mall*	0.51	11957	8	10189	4	n
L2_A	NS_1	real	position	cue-response	*fame*	*fortune*	0.487	5485	4	11815	7	n
L2_A	NS_1	real	position	cue-response	*tap*	*dance*	0.206	11279	3	21799	5	n
L2_A	NS_1	real	position	cue-response	*train*	*track*	0.32	23990	5	33788	5	n
L2_A	NS_1	real	position	cue-response	*department*	*store*	0.47	32252	10	56147	5	n
L2_A	NS_1	real	position	cue-response	*brand*	*name*	0.38	13103	5	127139	4	n
L2_A	NS_1	real	position	response-cue	*guide*	*tour*	0.2	20646	5	25781	4	n
L2_A	NS_1	real	position	response-cue	*juice*	*orange*	0.655	15388	5	9755	6	n
L2_A	NS_1	real	position	response-cue	*being*	*human*	0.34	21695	5	101224	5	n
L2_A	NS_1	real	form	affix	*loosen*	*loose*	0.007	11299	6	11299	5	aj
L2_A	NS_1	real	form	affix	*unite*	*union*	0.006	6391	5	22380	5	v
L2_A	NS_1	real	form	formal	*drift*	*draft*	0.007	8311	5	12684	5	n
L2_A	NS_1	real	form	formal	*take*	*thank*	0.006	670745	4	88574	5	v
L2_A	NS_1	real	form	formal	*proud*	*loud*	0.006	17841	5	10324	4	aj
L2_A	NS_1	real	form	formal	*deal*	*meal*	0.007	57462	4	21556	4	n

L2_A	NS_1	real	form	formal	*contract*	*contact*	0.007	30906	8	25218	7	n
L2_A	NS_1	real	form	formal	*sword*	*word*	0.006	7029	5	152891	4	n
L2_A	NS_1	real	form	two-step	*stake*	*sauce*	0.007	12603	5	15903	5	n
L2_A	NS_1	real	form	two-step	*chief*	*food*	0.006	26456	5	107728	4	n
L2_A	NS_1	real	unrelated	NA	*wagon*	*file*	NA	6577	5	20832	4	not controlled
L2_A	NS_1	real	unrelated	NA	*simple*	*fire*	NA	50583	6	59386	4	not controlled
L2_A	NS_1	real	unrelated	NA	*glass*	*thumb*	NA	49686	5	7659	5	not controlled
L2_A	NS_1	real	unrelated	NA	*complex*	*drugs*	NA	23751	7	86231	5	not controlled
L2_A	NS_1	real	unrelated	NA	*chair*	*throw*	NA	43256	5	57784	5	not controlled
L2_A	NS_1	real	unrelated	NA	*pen*	*flower*	NA	8117	3	25642	6	not controlled
L2_A	NS_1	real	unrelated	NA	*boiled*	*horse*	NA	6180	6	30993	5	not controlled
L2_A	NS_1	real	unrelated	NA	*storm*	*healthy*	NA	22562	5	26009	7	not controlled
L2_A	NS_1	real	unrelated	NA	*puzzle*	*brown*	NA	6153	6	21175	5	not controlled
L2_A	NS_1	real	unrelated	NA	*carry*	*escape*	NA	79513	5	17195	6	not controlled
L2_A	NS_1	real	unrelated	NA	*criticize*	*read*	NA	12906	9	114094	4	not controlled
L2_A	NS_1	real	unrelated	NA	*bill*	*plate*	NA	49011	4	24592	5	not controlled
L2_A	NS_1	real	unrelated	NA	*clothes*	*pass*	NA	27033	7	86184	4	not controlled
L2_A	NS_1	real	unrelated	NA	*ship*	*form*	NA	32588	4	78493	4	not controlled
L2_A	NS_1	real	unrelated	NA	*card*	*jeans*	NA	43605	4	8851	5	not controlled
L2_A	NS_1	real	unrelated	NA	*register*	*nice*	NA	9389	8	44792	4	not controlled
L2_A	NS_1	real	unrelated	NA	*sleeve*	*dog*	NA	5855	6	52347	3	not controlled
L2_A	NS_1	real	unrelated	NA	*log*	*trick*	NA	8606	3	10041	5	not controlled
L2_A	NS_1	real	unrelated	NA	*tooth*	*degree*	NA	20515	5	50612	6	not controlled
L2_A	NS_1	real	unrelated	NA	*onion*	*officer*	NA	13769	5	57617	7	not controlled
L2_A	NS_1	real	unrelated	NA	*building*	*care*	NA	78487	8	88862	4	not controlled
L2_A	NS_1	real	unrelated	NA	*surprise*	*can*	NA	22275	8	10718	3	not controlled
L2_A	NS_1	real	unrelated	NA	*smooth*	*tax*	NA	14148	6	80713	3	not controlled
L2_A	NS_1	real	unrelated	NA	*bottle*	*score*	NA	21569	6	39294	5	not controlled
L2_A	NS_1	real	unrelated	NA	*seed*	*race*	NA	19079	4	54838	4	not controlled
L2_A	NS_1	real	unrelated	NA	*voter*	*hair*	NA	27768	5	69564	4	not controlled
L2_A	NS_1	real	unrelated	NA	*security*	*diet*	NA	58914	8	16933	4	not controlled

L2_A	NS_1	real	unrelated	NA	*final*	*drain*	NA	43589	5	8123	5	not controlled
L2_A	NS_1	real	unrelated	NA	*green*	*winner*	NA	44673	5	19216	6	not controlled
L2_A	NS_1	real	unrelated	NA	*intent*	*tree*	NA	6879	6	66630	4	not controlled
L2_A	NS_1	pseudo	NA	NA	*sharp*	*lond*	NA	17403	5	NA	4	aj
L2_A	NS_1	pseudo	NA	NA	*dark*	*thirt*	NA	47565	4	NA	5	aj
L2_A	NS_1	pseudo	NA	NA	*glad*	*eltow*	NA	15556	4	NA	5	aj
L2_A	NS_1	pseudo	NA	NA	*public*	*gund*	NA	119825	6	NA	4	aj
L2_A	NS_1	pseudo	NA	NA	*crucial*	*haddle*	NA	14234	7	NA	6	aj
L2_A	NS_1	pseudo	NA	NA	*practical*	*cricky*	NA	15303	9	NA	6	aj
L2_A	NS_1	pseudo	NA	NA	*rich*	*twank*	NA	35940	4	NA	5	aj
L2_A	NS_1	pseudo	NA	NA	*tough*	*fow*	NA	39600	5	NA	3	aj
L2_A	NS_1	pseudo	NA	NA	*amazing*	*prudator*	NA	15124	7	NA	8	aj
L2_A	NS_1	pseudo	NA	NA	*yellow*	*rammary*	NA	22452	6	NA	7	aj
L2_A	NS_1	pseudo	NA	NA	*soil*	*aggraise*	NA	20916	4	NA	8	n
L2_A	NS_1	pseudo	NA	NA	*girl*	*corth*	NA	110409	4	NA	5	n
L2_A	NS_1	pseudo	NA	NA	*branch*	*wubs*	NA	19633	6	NA	4	n
L2_A	NS_1	pseudo	NA	NA	*sale*	*tam*	NA	15351	4	NA	3	n
L2_A	NS_1	pseudo	NA	NA	*skill*	*rog*	NA	50431	5	NA	3	n
L2_A	NS_1	pseudo	NA	NA	*pan*	*brifter*	NA	14148	3	NA	7	n
L2_A	NS_1	pseudo	NA	NA	*pack*	*almful*	NA	15169	4	NA	6	n
L2_A	NS_1	pseudo	NA	NA	*blanket*	*pedar*	NA	9385	7	NA	5	n
L2_A	NS_1	pseudo	NA	NA	*clothing*	*icolation*	NA	12078	8	NA	9	n
L2_A	NS_1	pseudo	NA	NA	*pose*	*scure*	NA	15230	4	NA	5	n
L2_A	NS_1	pseudo	NA	NA	*whale*	*blucky*	NA	6928	5	NA	6	n
L2_A	NS_1	pseudo	NA	NA	*shoe*	*penseless*	NA	26945	4	NA	9	n
L2_A	NS_1	pseudo	NA	NA	*ski*	*daping*	NA	11656	3	NA	6	n
L2_A	NS_1	pseudo	NA	NA	*father*	*spon*	NA	145051	6	NA	4	n
L2_A	NS_1	pseudo	NA	NA	*box*	*sauto*	NA	49667	3	NA	5	n
L2_A	NS_1	pseudo	NA	NA	*fiction*	*tespair*	NA	11701	7	NA	7	n
L2_A	NS_1	pseudo	NA	NA	*city*	*sefund*	NA	132684	4	NA	6	n
L2_A	NS_1	pseudo	NA	NA	*cage*	*repail*	NA	6621	4	NA	6	n

L2_A	NS_1	pseudo	NA	NA	*knife*	*fugs*	NA	15792	5	NA	4	n
L2_A	NS_1	pseudo	NA	NA	*market*	*pubby*	NA	100435	6	NA	5	n
L2_A	NS_1	pseudo	NA	NA	*suicide*	*geavens*	NA	15145	7	NA	7	n
L2_A	NS_1	pseudo	NA	NA	*team*	*beeg*	NA	131489	4	NA	4	n
L2_A	NS_1	pseudo	NA	NA	*toilet*	*bingside*	NA	6939	6	NA	8	n
L2_A	NS_1	pseudo	NA	NA	*pig*	*drair*	NA	8307	3	NA	5	n
L2_A	NS_1	pseudo	NA	NA	*master*	*goncho*	NA	18880	6	NA	6	n
L2_A	NS_1	pseudo	NA	NA	*pocket*	*sanner*	NA	23580	6	NA	6	n
L2_A	NS_1	pseudo	NA	NA	*fist*	*slea*	NA	7729	4	NA	4	n
L2_A	NS_1	pseudo	NA	NA	*boot*	*zate*	NA	15033	4	NA	4	n
L2_A	NS_1	pseudo	NA	NA	*cheese*	*fervo*	NA	17416	6	NA	5	n
L2_A	NS_1	pseudo	NA	NA	*fan*	*drinkle*	NA	32919	3	NA	7	n
L2_A	NS_1	pseudo	NA	NA	*strength*	*feven*	NA	29769	8	NA	5	n
L2_A	NS_1	pseudo	NA	NA	*ladder*	*cleed*	NA	5720	6	NA	5	n
L2_A	NS_1	pseudo	NA	NA	*weed*	*doise*	NA	5642	4	NA	5	n
L2_A	NS_1	pseudo	NA	NA	*pound*	*neep*	NA	29946	5	NA	4	n
L2_A	NS_1	pseudo	NA	NA	*star*	*mot*	NA	73695	4	NA	3	n
L2_A	NS_1	pseudo	NA	NA	*level*	*coster*	NA	121704	5	NA	6	n
L2_A	NS_1	pseudo	NA	NA	*meat*	*disoble*	NA	20271	4	NA	7	n
L2_A	NS_1	pseudo	NA	NA	*brain*	*shorus*	NA	32852	5	NA	6	n
L2_A	NS_1	pseudo	NA	NA	*person*	*squing*	NA	113650	6	NA	6	n
L2_A	NS_1	pseudo	NA	NA	*chart*	*thamp*	NA	11861	5	NA	5	n
L2_A	NS_1	pseudo	NA	NA	*walk*	*joeful*	NA	113787	4	NA	6	v
L2_A	NS_1	pseudo	NA	NA	*limit*	*redge*	NA	28901	5	NA	5	v
L2_A	NS_1	pseudo	NA	NA	*write*	*dathway*	NA	161824	5	NA	7	v
L2_A	NS_1	pseudo	NA	NA	*shift*	*irpact*	NA	21323	5	NA	6	v
L2_A	NS_1	pseudo	NA	NA	*breathe*	*nery*	NA	15813	7	NA	4	v
L2_A	NS_1	pseudo	NA	NA	*win*	*sall*	NA	111478	3	NA	4	v
L2_A	NS_1	pseudo	NA	NA	*pour*	*slein*	NA	19300	4	NA	5	v
L2_A	NS_1	pseudo	NA	NA	*force*	*procks*	NA	108005	5	NA	6	v
L2_A	NS_1	pseudo	NA	NA	*offer*	*hoag*	NA	106473	5	NA	4	v

L2_A	NS_1	pseudo	NA	NA	*elect*	*cayward*	NA	15350	5	NA	7	v
L2_B	NS_1	real	meaning	antonym	*north*	*south*	0.77	68046	5	60630	5	aj
L2_B	NS_1	real	meaning	antonym	*buy*	*sell*	0.35	101105	3	87865	4	v
L2_B	NS_1	real	meaning	conceptual other	*wisdom*	*smart*	0.21	10561	6	19370	5	aj
L2_B	NS_1	real	meaning	conceptual other	*sky*	*cloud*	0.22	35141	3	19214	5	n
L2_B	NS_1	real	meaning	coordination	*vegetable*	*fruit*	0.22	19363	9	22401	5	n
L2_B	NS_1	real	meaning	coordination	*van*	*truck*	0.214	8364	3	31536	5	n
L2_B	NS_1	real	meaning	hierarchical set	*transportation*	*car*	0.595	14516	14	133571	3	n
L2_B	NS_1	real	meaning	hierarchical set	*cousin*	*relative*	0.24	12155	6	13541	8	n
L2_B	NS_1	real	meaning	synonym	*stone*	*rock*	0.629	32531	5	45225	4	n
L2_B	NS_1	real	meaning	synonym	*reach*	*grab*	0.255	7982	5	26951	4	v
L2_B	NS_1	real	position	cue-response	*alarm*	*clock*	0.388	7631	5	12395	5	n
L2_B	NS_1	real	position	cue-response	*spare*	*tire*	0.41	5492	5	11494	4	n
L2_B	NS_1	real	position	cue-response	*fishing*	*pole*	0.2	18103	7	11446	4	n
L2_B	NS_1	real	position	cue-response	*mouth*	*wash*	0.201	40200	5	16314	4	n
L2_B	NS_1	real	position	cue-response	*jet*	*plane*	0.662	9690	3	33900	5	n
L2_B	NS_1	real	position	cue-response	*border*	*line*	0.206	28636	6	135986	4	n
L2_B	NS_1	real	position	cue-response	*abstract*	*art*	0.23	7165	8	117851	3	n
L2_B	NS_1	real	position	response-cue	*core*	*apple*	0.3	18571	4	12172	5	n
L2_B	NS_1	real	position	response-cue	*album*	*record*	0.33	15869	5	77509	6	n
L2_B	NS_1	real	position	response-cue	*link*	*chain*	0.4	21335	4	19688	5	n
L2_B	NS_1	real	form	affix	*respond*	*response*	0.007	42139	7	56342	8	n
L2_B	NS_1	real	form	affix	*solve*	*resolve*	0.007	19501	5	15260	7	v
L2_B	NS_1	real	form	formal	*chance*	*change*	0.005	62682	6	123183	6	v
L2_B	NS_1	real	form	formal	*decrease*	*peace*	0.007	9291	8	42273	5	n
L2_B	NS_1	real	form	formal	*third*	*tired*	0.007	67037	5	18597	5	aj
L2_B	NS_1	real	form	formal	*shove*	*oven*	0.006	6327	5	11454	4	n
L2_B	NS_1	real	form	formal	*deer*	*fear*	0.007	11665	4	38857	4	n
L2_B	NS_1	real	form	formal	*mix*	*fix*	0.008	20642	3	19349	3	n
L2_B	NS_1	real	form	two-step	*warn*	*cold*	0.006	19996	4	44649	4	aj
L2_B	NS_1	real	form	two-step	*content*	*land*	0.007	21821	7	69750	4	n

L2_B	NS_1	real	unrelated	NA	*sick*	*pencil*	NA	20906	4	5936	6	not controlled
L2_B	NS_1	real	unrelated	NA	*catch*	*rough*	NA	68214	5	12365	5	not controlled
L2_B	NS_1	real	unrelated	NA	*illegal*	*hot*	NA	17258	7	54601	3	not controlled
L2_B	NS_1	real	unrelated	NA	*shirt*	*table*	NA	21486	5	75228	5	not controlled
L2_B	NS_1	real	unrelated	NA	*window*	*beard*	NA	27917	6	5528	5	not controlled
L2_B	NS_1	real	unrelated	NA	*rose*	*fail*	NA	8988	4	47503	4	not controlled
L2_B	NS_1	real	unrelated	NA	*finger*	*blue*	NA	40842	6	47622	4	not controlled
L2_B	NS_1	real	unrelated	NA	*flee*	*hard*	NA	10508	4	86817	4	not controlled
L2_B	NS_1	real	unrelated	NA	*dish*	*haul*	NA	18887	4	6401	4	not controlled
L2_B	NS_1	real	unrelated	NA	*aluminum*	*kind*	NA	6439	8	155032	4	not controlled
L2_B	NS_1	real	unrelated	NA	*reader*	*party*	NA	6468	8	112962	5	not controlled
L2_B	NS_1	real	unrelated	NA	*critic*	*phone*	NA	28244	6	71599	5	not controlled
L2_B	NS_1	real	unrelated	NA	*wheel*	*piece*	NA	13688	5	68901	5	not controlled
L2_B	NS_1	real	unrelated	NA	*hurricane*	*credit*	NA	31442	6	34578	6	not controlled
L2_B	NS_1	real	unrelated	NA	*treat*	*cruise*	NA	40264	5	5434	6	not controlled
L2_B	NS_1	real	unrelated	NA	*animal*	*income*	NA	53127	6	34925	6	not controlled
L2_B	NS_1	real	unrelated	NA	*cabinet*	*black*	NA	27917	7	150718	5	not controlled
L2_B	NS_1	real	unrelated	NA	*designer*	*blocks*	NA	14210	8	13688	6	not controlled
L2_B	NS_1	real	unrelated	NA	*secure*	*brush*	NA	13181	6	9806	5	not controlled
L2_B	NS_1	real	unrelated	NA	*soccer*	*cabin*	NA	9212	6	13688	5	not controlled
L2_B	NS_1	real	unrelated	NA	*eggs*	*vote*	NA	27917	4	39464	4	not controlled
L2_B	NS_1	real	unrelated	NA	*intense*	*die*	NA	27917	7	98376	3	not controlled
L2_B	NS_1	real	unrelated	NA	*cash*	*sample*	NA	21343	4	32436	6	not controlled
L2_B	NS_1	real	unrelated	NA	*police*	*bottom*	NA	85880	6	24653	6	not controlled
L2_B	NS_1	real	unrelated	NA	*define*	*leap*	NA	34501	7	12891	4	not controlled
L2_B	NS_1	real	unrelated	NA	*opinion*	*pace*	NA	33958	6	8376	4	not controlled
L2_B	NS_1	real	unrelated	NA	*rain*	*share*	NA	24134	4	54010	5	not controlled
L2_B	NS_1	real	unrelated	NA	*four*	*ride*	NA	150646	4	30476	4	not controlled
L2_B	NS_1	real	unrelated	NA	*firm*	*mayor*	NA	44704	4	13903	5	not controlled
L2_B	NS_1	real	unrelated	NA	*season*	*feed*	NA	83743	6	28494	4	not controlled
L2_B	NS_1	pseudo	NA	NA	*important*	*grike*	NA	144194	9	NA	5	aj

L2_B	NS_1	pseudo	NA	NA	*lucky*	*besh*	NA	16550	5	NA	4	aj
L2_B	NS_1	pseudo	NA	NA	*visual*	*fleak*	NA	17316	6	NA	5	aj
L2_B	NS_1	pseudo	NA	NA	*short*	*hather*	NA	60451	5	NA	6	aj
L2_B	NS_1	pseudo	NA	NA	*bright*	*theer*	NA	29780	6	NA	5	aj
L2_B	NS_1	pseudo	NA	NA	*perfect*	*teasant*	NA	33456	7	NA	7	aj
L2_B	NS_1	pseudo	NA	NA	*low*	*vaming*	NA	108990	3	NA	6	aj
L2_B	NS_1	pseudo	NA	NA	*dirty*	*fraceable*	NA	11112	5	NA	9	aj
L2_B	NS_1	pseudo	NA	NA	*pink*	*affitude*	NA	13849	4	NA	8	aj
L2_B	NS_1	pseudo	NA	NA	*mild*	*vutt*	NA	7211	4	NA	4	aj
L2_B	NS_1	pseudo	NA	NA	*gang*	*ruge*	NA	12662	4	NA	4	n
L2_B	NS_1	pseudo	NA	NA	*research*	*churks*	NA	114802	8	NA	6	n
L2_B	NS_1	pseudo	NA	NA	*bowl*	*ostane*	NA	20662	4	NA	6	n
L2_B	NS_1	pseudo	NA	NA	*summer*	*tocoa*	NA	62503	6	NA	5	n
L2_B	NS_1	pseudo	NA	NA	*mouse*	*sponto*	NA	9449	5	NA	6	n
L2_B	NS_1	pseudo	NA	NA	*chest*	*roneless*	NA	22508	5	NA	8	n
L2_B	NS_1	pseudo	NA	NA	*eye*	*decord*	NA	169150	3	NA	6	n
L2_B	NS_1	pseudo	NA	NA	*arm*	*yick*	NA	84865	3	NA	4	n
L2_B	NS_1	pseudo	NA	NA	*teacher*	*steezen*	NA	116100	7	NA	7	n
L2_B	NS_1	pseudo	NA	NA	*morning*	*grova*	NA	114002	7	NA	5	n
L2_B	NS_1	pseudo	NA	NA	*childhood*	*serm*	NA	16268	9	NA	4	n
L2_B	NS_1	pseudo	NA	NA	*service*	*drivus*	NA	146122	7	NA	6	n
L2_B	NS_1	pseudo	NA	NA	*tale*	*loe*	NA	15004	4	NA	3	n
L2_B	NS_1	pseudo	NA	NA	*plant*	*dit*	NA	63476	5	NA	3	n
L2_B	NS_1	pseudo	NA	NA	*job*	*boet*	NA	154743	3	NA	4	n
L2_B	NS_1	pseudo	NA	NA	*sun*	*purmise*	NA	32646	3	NA	7	n
L2_B	NS_1	pseudo	NA	NA	*rope*	*nar*	NA	9300	4	NA	3	n
L2_B	NS_1	pseudo	NA	NA	*grade*	*frot*	NA	27178	5	NA	4	n
L2_B	NS_1	pseudo	NA	NA	*rabbit*	*tinking*	NA	6095	6	NA	7	n
L2_B	NS_1	pseudo	NA	NA	*scholar*	*honal*	NA	17482	7	NA	5	n
L2_B	NS_1	pseudo	NA	NA	*gene*	*jeist*	NA	15377	4	NA	5	n

L2_B	NS_1	pseudo	NA	NA	*boy*	*trumble*	NA	107447	3	NA	7	n
L2_B	NS_1	pseudo	NA	NA	*goat*	*slound*	NA	5459	4	NA	6	n
L2_B	NS_1	pseudo	NA	NA	*reason*	*eptol*	NA	106863	6	NA	5	n
L2_B	NS_1	pseudo	NA	NA	*pitcher*	*lervid*	NA	8835	7	NA	6	n
L2_B	NS_1	pseudo	NA	NA	*fact*	*zaly*	NA	164401	4	NA	4	n
L2_B	NS_1	pseudo	NA	NA	*office*	*farsh*	NA	114791	6	NA	5	n
L2_B	NS_1	pseudo	NA	NA	*result*	*pindless*	NA	116277	6	NA	8	n
L2_B	NS_1	pseudo	NA	NA	*blood*	*peedbag*	NA	56351	5	NA	7	n
L2_B	NS_1	pseudo	NA	NA	*education*	*slout*	NA	113731	9	NA	5	n
L2_B	NS_1	pseudo	NA	NA	*water*	*pid*	NA	167666	5	NA	3	n
L2_B	NS_1	pseudo	NA	NA	*news*	*pum*	NA	70051	4	NA	3	n
L2_B	NS_1	pseudo	NA	NA	*midnight*	*droe*	NA	8440	8	NA	4	n
L2_B	NS_1	pseudo	NA	NA	*parking*	*cigh*	NA	14970	7	NA	4	n
L2_B	NS_1	pseudo	NA	NA	*lung*	*parg*	NA	9206	4	NA	4	n
L2_B	NS_1	pseudo	NA	NA	*growth*	*teps*	NA	50904	6	NA	4	n
L2_B	NS_1	pseudo	NA	NA	*friend*	*biser*	NA	142697	6	NA	5	n
L2_B	NS_1	pseudo	NA	NA	*sock*	*madnoss*	NA	6273	4	NA	7	n
L2_B	NS_1	pseudo	NA	NA	*passion*	*gaws*	NA	14632	7	NA	4	n
L2_B	NS_1	pseudo	NA	NA	*driver*	*ranish*	NA	31633	6	NA	6	n
L2_B	NS_1	pseudo	NA	NA	*exist*	*spratch*	NA	39341	5	NA	7	v
L2_B	NS_1	pseudo	NA	NA	*create*	*soal*	NA	119419	6	NA	4	v
L2_B	NS_1	pseudo	NA	NA	*impose*	*kaws*	NA	14881	6	NA	4	v
L2_B	NS_1	pseudo	NA	NA	*edit*	*striggle*	NA	6157	4	NA	8	v
L2_B	NS_1	pseudo	NA	NA	*accept*	*heptile*	NA	49952	6	NA	7	v
L2_B	NS_1	pseudo	NA	NA	*sit*	*defient*	NA	147185	3	NA	7	v
L2_B	NS_1	pseudo	NA	NA	*touch*	*carm*	NA	34737	5	NA	4	v
L2_B	NS_1	pseudo	NA	NA	*ruin*	*sisked*	NA	6267	4	NA	6	v
L2_B	NS_1	pseudo	NA	NA	*listen*	*paith*	NA	64984	6	NA	5	v
L2_B	NS_1	pseudo	NA	NA	*enable*	*gight*	NA	16293	6	NA	5	v

L2_C	NS_2	real	meaning	antonym	*fail*	*pass*	0.21	47503	4	86184	4	v
L2_C	NS_2	real	meaning	antonym	*rough*	*smooth*	0.35	12365	5	14148	6	aj
L2_C	NS_2	real	meaning	conceptual other	*window*	*glass*	0.256	68303	6	49686	5	n
L2_C	NS_2	real	meaning	conceptual other	*beard*	*hair*	0.2	5528	5	69564	4	n
L2_C	NS_2	real	meaning	coordination	*table*	*chair*	0.75	75228	5	43256	5	n
L2_C	NS_2	real	meaning	coordination	*pencil*	*pen*	0.47	5936	6	8117	3	n
L2_C	NS_2	real	meaning	hierarchical set	*animal*	*dog*	0.293	53127	6	52347	3	n
L2_C	NS_2	real	meaning	hierarchical set	*hurricane*	*storm*	0.22	6439	9	22562	5	n
L2_C	NS_2	real	meaning	synonym	*kind*	*nice*	0.36	155032	4	44792	4	aj
L2_C	NS_2	real	meaning	synonym	*haul*	*carry*	0.226	6401	4	79513	5	v
L2_C	NS_2	real	position	cue-response	*aluminum*	*can*	0.315	6468	8	10718	3	n
L2_C	NS_2	real	position	cue-response	*phone*	*bill*	0.207	71599	5	49011	3	n
L2_C	NS_2	real	position	cue-response	*designer*	*clothes*	0.427	14210	8	27033	7	n
L2_C	NS_2	real	position	cue-response	*cruise*	*ship*	0.44	5434	6	32588	4	n
L2_C	NS_2	real	position	cue-response	*credit*	*card*	0.65	34578	6	43605	4	n
L2_C	NS_2	real	position	cue-response	*income*	*tax*	0.22	34925	6	80713	3	n
L2_C	NS_2	real	position	cue-response	*illegal*	*drugs*	0.273	17258	7	86231	5	n
L2_C	NS_2	real	position	response-cue	*cabin*	*log*	0.415	9955	5	8606	3	n
L2_C	NS_2	real	position	response-cue	*brush*	*tooth*	0.277	9806	5	20515	5	n
L2_C	NS_2	real	position	response-cue	*blocks*	*building*	0.253	28826	6	78487	8	n
L2_C	NS_2	real	form	affix	*vote*	*voter*	0.006	39464	4	27768	5	n
L2_C	NS_2	real	form	affix	*reader*	*read*	0.007	31442	6	114094	4	v
L2_C	NS_2	real	form	formal	*bottom*	*bottle*	0.007	24653	6	21569	6	n
L2_C	NS_2	real	form	formal	*sample*	*simple*	0.007	32436	6	50583	6	aj
L2_C	NS_2	real	form	formal	*opinion*	*onion*	0.005	34501	7	13769	5	n
L2_C	NS_2	real	form	formal	*feed*	*seed*	0.008	28494	4	19079	4	n
L2_C	NS_2	real	form	formal	*share*	*care*	0.007	54010	5	42978	4	v
L2_C	NS_2	real	form	formal	*define*	*final*	0.007	33958	6	43589	5	aj
L2_C	NS_2	real	form	two-step	*leap*	*green*	0.005	8376	4	44673	5	n
L2_C	NS_2	real	form	two-step	*mayor*	*degree*	0.007	13903	5	50612	6	n
L2_C	NS_2	real	unrelated	NA	*late*	*south*	NA	86421	4	60630	5	not controlled

L2_C	NS_2	real	unrelated	NA	*exit*	*fruit*	NA	7333	4	22401	5	not controlled
L2_C	NS_2	real	unrelated	NA	*bee*	*sell*	NA	6342	3	87865	4	not controlled
L2_C	NS_2	real	unrelated	NA	*calendar*	*smart*	NA	5258	8	19370	5	not controlled
L2_C	NS_2	real	unrelated	NA	*square*	*car*	NA	11630	6	133571	3	not controlled
L2_C	NS_2	real	unrelated	NA	*spoon*	*truck*	NA	6194	5	31536	5	not controlled
L2_C	NS_2	real	unrelated	NA	*color*	*relative*	NA	56978	5	13541	8	not controlled
L2_C	NS_2	real	unrelated	NA	*robot*	*cloud*	NA	6023	5	19214	5	not controlled
L2_C	NS_2	real	unrelated	NA	*dumb*	*grab*	NA	5486	4	26951	4	not controlled
L2_C	NS_2	real	unrelated	NA	*trim*	*rock*	NA	5113	4	45225	4	not controlled
L2_C	NS_2	real	unrelated	NA	*tissue*	*clock*	NA	6180	6	12395	5	not controlled
L2_C	NS_2	real	unrelated	NA	*shopping*	*pole*	NA	11957	8	11446	4	not controlled
L2_C	NS_2	real	unrelated	NA	*fame*	*tire*	NA	5485	4	11494	4	not controlled
L2_C	NS_2	real	unrelated	NA	*tap*	*plane*	NA	11279	3	33900	5	not controlled
L2_C	NS_2	real	unrelated	NA	*train*	*wash*	NA	23990	5	16314	4	not controlled
L2_C	NS_2	real	unrelated	NA	*department*	*apple*	NA	32252	10	12172	5	not controlled
L2_C	NS_2	real	unrelated	NA	*brand*	*response*	NA	13103	5	56342	8	not controlled
L2_C	NS_2	real	unrelated	NA	*guide*	*resolve*	NA	20646	5	15260	7	not controlled
L2_C	NS_2	real	unrelated	NA	*juice*	*record*	NA	15388	5	77509	6	not controlled
L2_C	NS_2	real	unrelated	NA	*being*	*chain*	NA	21695	5	19688	5	not controlled
L2_C	NS_2	real	unrelated	NA	*loosen*	*art*	NA	11299	6	117851	3	not controlled
L2_C	NS_2	real	unrelated	NA	*unite*	*tired*	NA	11299	5	18597	5	not controlled
L2_C	NS_2	real	unrelated	NA	*drift*	*peace*	NA	8311	5	42273	5	not controlled
L2_C	NS_2	real	unrelated	NA	*take*	*line*	NA	670745	4	135986	4	not controlled
L2_C	NS_2	real	unrelated	NA	*deal*	*oven*	NA	57462	4	11454	4	not controlled
L2_C	NS_2	real	unrelated	NA	*proud*	*fix*	NA	17841	5	19349	3	not controlled
L2_C	NS_2	real	unrelated	NA	*contract*	*fear*	NA	30906	8	38857	4	not controlled
L2_C	NS_2	real	unrelated	NA	*stake*	*change*	NA	12603	5	123183	6	not controlled
L2_C	NS_2	real	unrelated	NA	*chief*	*cold*	NA	26456	5	44649	4	not controlled
L2_C	NS_2	real	unrelated	NA	*sword*	*land*	NA	7029	5	69750	4	not controlled
L2_C	NS_2	pseudo	NA	NA	*ordinary*	*soble*	NA	14776	8	NA	5	aj
L2_C	NS_2	pseudo	NA	NA	*empty*	*tarpal*	NA	23365	5	NA	6	aj

L2_C	NS_2	pseudo	NA	NA	*wet*	*mide*	NA	13963	3	NA	4	aj
L2_C	NS_2	pseudo	NA	NA	*oak*	*gustle*	NA	5938	3	NA	6	aj
L2_C	NS_2	pseudo	NA	NA	*grand*	*bly*	NA	15659	5	NA	3	aj
L2_C	NS_2	pseudo	NA	NA	*best*	*purf*	NA	124850	4	NA	4	aj
L2_C	NS_2	pseudo	NA	NA	*silent*	*pash*	NA	16802	6	NA	4	aj
L2_C	NS_2	pseudo	NA	NA	*helpful*	*bup*	NA	10120	7	NA	3	aj
L2_C	NS_2	pseudo	NA	NA	*angry*	*hesignate*	NA	21485	5	NA	9	aj
L2_C	NS_2	pseudo	NA	NA	*young*	*mip*	NA	160011	5	NA	3	aj
L2_C	NS_2	pseudo	NA	NA	*index*	*trull*	NA	12751	5	NA	5	n
L2_C	NS_2	pseudo	NA	NA	*bridge*	*defirm*	NA	21497	6	NA	6	n
L2_C	NS_2	pseudo	NA	NA	*jail*	*tup*	NA	13324	4	NA	3	n
L2_C	NS_2	pseudo	NA	NA	*vision*	*kint*	NA	32358	6	NA	4	n
L2_C	NS_2	pseudo	NA	NA	*region*	*dape*	NA	50914	6	NA	4	n
L2_C	NS_2	pseudo	NA	NA	*inspector*	*troad*	NA	8652	9	NA	5	n
L2_C	NS_2	pseudo	NA	NA	*body*	*dobs*	NA	125165	4	NA	4	n
L2_C	NS_2	pseudo	NA	NA	*carbon*	*slutter*	NA	10799	6	NA	7	n
L2_C	NS_2	pseudo	NA	NA	*minute*	*beflect*	NA	126660	6	NA	6	n
L2_C	NS_2	pseudo	NA	NA	*variety*	*hond*	NA	34242	7	NA	4	n
L2_C	NS_2	pseudo	NA	NA	*worker*	*gauterize*	NA	69962	6	NA	9	n
L2_C	NS_2	pseudo	NA	NA	*pizza*	*shenic*	NA	7130	5	NA	6	n
L2_C	NS_2	pseudo	NA	NA	*threat*	*fimple*	NA	37022	6	NA	6	n
L2_C	NS_2	pseudo	NA	NA	*ice*	*hiaper*	NA	31686	3	NA	6	n
L2_C	NS_2	pseudo	NA	NA	*web*	*ratter*	NA	6411	3	NA	6	n
L2_C	NS_2	pseudo	NA	NA	*focus*	*crip*	NA	57177	5	NA	4	n
L2_C	NS_2	pseudo	NA	NA	*iron*	*uldone*	NA	15043	4	NA	6	n
L2_C	NS_2	pseudo	NA	NA	*skin*	*dism*	NA	39893	4	NA	4	n
L2_C	NS_2	pseudo	NA	NA	*air*	*gry*	NA	105932	3	NA	3	n
L2_C	NS_2	pseudo	NA	NA	*town*	*roft*	NA	79821	4	NA	4	n
L2_C	NS_2	pseudo	NA	NA	*page*	*criad*	NA	55937	4	NA	5	n
L2_C	NS_2	pseudo	NA	NA	*boat*	*larmish*	NA	32079	4	NA	7	n
L2_C	NS_2	pseudo	NA	NA	*injury*	*nerits*	NA	23935	6	NA	6	n

L2_C	NS_2	pseudo	NA	NA	*sheet*	*hesk*	NA	20787	5	NA	4	n
L2_C	NS_2	pseudo	NA	NA	*poet*	*fote*	NA	10840	4	NA	4	n
L2_C	NS_2	pseudo	NA	NA	*clue*	*thrilly*	NA	8732	4	NA	7	n
L2_C	NS_2	pseudo	NA	NA	*warning*	*ralt*	NA	16205	7	NA	4	n
L2_C	NS_2	pseudo	NA	NA	*title*	*menizen*	NA	29210	5	NA	7	n
L2_C	NS_2	pseudo	NA	NA	*jacket*	*tenny*	NA	15692	6	NA	5	n
L2_C	NS_2	pseudo	NA	NA	*adviser*	*soose*	NA	12112	7	NA	5	n
L2_C	NS_2	pseudo	NA	NA	*crisis*	*theil*	NA	32924	6	NA	5	n
L2_C	NS_2	pseudo	NA	NA	*war*	*cise*	NA	117804	3	NA	4	n
L2_C	NS_2	pseudo	NA	NA	*style*	*flad*	NA	40889	5	NA	4	n
L2_C	NS_2	pseudo	NA	NA	*staff*	*kig*	NA	50177	5	NA	3	n
L2_C	NS_2	pseudo	NA	NA	*justice*	*rorder*	NA	25377	7	NA	6	n
L2_C	NS_2	pseudo	NA	NA	*spirit*	*berdict*	NA	32942	6	NA	7	n
L2_C	NS_2	pseudo	NA	NA	*area*	*glarm*	NA	165812	4	NA	5	n
L2_C	NS_2	pseudo	NA	NA	*health*	*naving*	NA	117762	6	NA	6	n
L2_C	NS_2	pseudo	NA	NA	*stream*	*prack*	NA	15400	6	NA	5	n
L2_C	NS_2	pseudo	NA	NA	*payment*	*revoid*	NA	16704	7	NA	6	n
L2_C	NS_2	pseudo	NA	NA	*skip*	*dass*	NA	5635	4	NA	4	v
L2_C	NS_2	pseudo	NA	NA	*blend*	*ugload*	NA	7238	5	NA	6	v
L2_C	NS_2	pseudo	NA	NA	*provide*	*kip*	NA	150879	7	NA	3	v
L2_C	NS_2	pseudo	NA	NA	*adopt*	*darf*	NA	22880	5	NA	4	v
L2_C	NS_2	pseudo	NA	NA	*assist*	*hospel*	NA	13748	6	NA	6	v
L2_C	NS_2	pseudo	NA	NA	*spend*	*vead*	NA	114569	5	NA	4	v
L2_C	NS_2	pseudo	NA	NA	*volunteer*	*iliotic*	NA	14958	9	NA	7	v
L2_C	NS_2	pseudo	NA	NA	*include*	*brudge*	NA	133563	7	NA	6	v
L2_C	NS_2	pseudo	NA	NA	*speak*	*gie*	NA	117358	5	NA	3	v
L2_C	NS_2	pseudo	NA	NA	*arrive*	*dumid*	NA	47435	6	NA	5	v
L2_D	NS_2	real	meaning	antonym	*healthy*	*sick*	0.23	26009	6	20906	4	aj
L2_D	NS_2	real	meaning	antonym	*throw*	*catch*	0.23	57784	5	68214	5	v
L2_D	NS_2	real	meaning	conceptual other	*fire*	*hot*	0.285	59386	4	54601	3	aj
L2_D	NS_2	real	meaning	conceptual other	*sleeve*	*shirt*	0.461	5855	6	21486	5	n

L2_D	NS_2	real	meaning	coordination	*brown*	*black*	0.338	21175	5	150718	5	n
L2_D	NS_2	real	meaning	coordination	*thumb*	*finger*	0.257	7659	5	40842	6	n
L2_D	NS_2	real	meaning	hierarchical set	*flower*	*rose*	0.248	25642	6	8988	4	n
L2_D	NS_2	real	meaning	hierarchical set	*plate*	*dish*	0.23	24592	5	18887	4	n
L2_D	NS_2	real	meaning	synonym	*escape*	*flee*	0.25	6358	6	10508	4	v
L2_D	NS_2	real	meaning	synonym	*complex*	*hard*	0.24	23030	5	28344	4	n
L2_D	NS_2	real	position	cue-response	*surprise*	*party*	0.24	22275	8	112962	5	n
L2_D	NS_2	real	position	cue-response	*file*	*cabinet*	0.24	20832	4	11846	7	n
L2_D	NS_2	real	position	cue-response	*wagon*	*wheel*	0.25	6577	5	18296	5	n
L2_D	NS_2	real	position	cue-response	*horse*	*ride*	0.261	30993	5	30476	4	n
L2_D	NS_2	real	position	cue-response	*trick*	*treat*	0.33	10041	5	40264	5	n
L2_D	NS_2	real	position	cue-response	*puzzle*	*piece*	0.235	6153	6	68901	5	n
L2_D	NS_2	real	position	cue-response	*boiled*	*eggs*	0.374	6180	6	27917	4	n
L2_D	NS_2	real	position	response-cue	*jeans*	*blue*	0.257	8851	5	47622	4	n
L2_D	NS_2	real	position	response-cue	*register*	*cash*	0.27	9389	8	21343	4	n
L2_D	NS_2	real	position	response-cue	*officer*	*police*	0.46	57617	7	85880	6	n
L2_D	NS_2	real	form	affix	*security*	*secure*	0.008	58914	8	13181	6	n
L2_D	NS_2	real	form	affix	*criticize*	*critic*	0.005	12906	9	28244	6	n
L2_D	NS_2	real	form	formal	*intent*	*intense*	0.006	6879	6	14452	7	aj
L2_D	NS_2	real	form	formal	*score*	*soccer*	0.007	39294	5	9212	6	n
L2_D	NS_2	real	form	formal	*race*	*pace*	0.007	54838	4	12891	4	n
L2_D	NS_2	real	form	formal	*diet*	*die*	0.007	16933	4	98376	3	v
L2_D	NS_2	real	form	formal	*drain*	*rain*	0.007	8123	5	24134	4	v
L2_D	NS_2	real	form	formal	*form*	*firm*	0.006	78493	4	44704	4	aj
L2_D	NS_2	real	form	two-step	*tree*	*four*	0.007	66630	4	150646	4	n
L2_D	NS_2	real	form	two-step	*winner*	*season*	0.007	19216	6	83743	6	n
L2_D	NS_2	real	unrelated	NA	*north*	*enter*	NA	68046	5	54479	5	not controlled
L2_D	NS_2	real	unrelated	NA	*buy*	*date*	NA	101105	3	31467	4	not controlled
L2_D	NS_2	real	unrelated	NA	*wisdom*	*honey*	NA	10561	6	11009	5	not controlled
L2_D	NS_2	real	unrelated	NA	*sky*	*circle*	NA	35141	3	24735	6	not controlled
L2_D	NS_2	real	unrelated	NA	*vegetable*	*early*	NA	19363	9	108171	5	not controlled

L2_D	NS_2	real	unrelated	NA	*van*	*fork*	NA	8364	3	7742	4	not controlled
L2_D	NS_2	real	unrelated	NA	*transportation*	*stupid*	NA	14516	14	12467	6	not controlled
L2_D	NS_2	real	unrelated	NA	*cousin*	*machine*	NA	12155	6	38407	7	not controlled
L2_D	NS_2	real	unrelated	NA	*stone*	*dance*	NA	32531	5	21799	5	not controlled
L2_D	NS_2	real	unrelated	NA	*reach*	*cut*	NA	92375	5	96012	3	not controlled
L2_D	NS_2	real	unrelated	NA	*alarm*	*paper*	NA	7631	5	75383	5	not controlled
L2_D	NS_2	real	unrelated	NA	*spare*	*mall*	NA	5492	5	10189	4	not controlled
L2_D	NS_2	real	unrelated	NA	*fishing*	*red*	NA	18103	7	66217	3	not controlled
L2_D	NS_2	real	unrelated	NA	*mouth*	*fortune*	NA	40200	5	11815	7	not controlled
L2_D	NS_2	real	unrelated	NA	*jet*	*store*	NA	9690	3	56147	5	not controlled
L2_D	NS_2	real	unrelated	NA	*border*	*name*	NA	28636	6	127139	4	not controlled
L2_D	NS_2	real	unrelated	NA	*abstract*	*tour*	NA	7165	8	25781	4	not controlled
L2_D	NS_2	real	unrelated	NA	*core*	*track*	NA	18571	4	33788	5	not controlled
L2_D	NS_2	real	unrelated	NA	*album*	*orange*	NA	15869	5	9755	6	not controlled
L2_D	NS_2	real	unrelated	NA	*link*	*meal*	NA	21335	4	21556	4	not controlled
L2_D	NS_2	real	unrelated	NA	*respond*	*loose*	NA	42139	7	11299	5	not controlled
L2_D	NS_2	real	unrelated	NA	*solve*	*draft*	NA	19501	5	12684	5	not controlled
L2_D	NS_2	real	unrelated	NA	*chance*	*union*	NA	62682	6	22380	5	not controlled
L2_D	NS_2	real	unrelated	NA	*decrease*	*word*	NA	9291	8	152891	4	not controlled
L2_D	NS_2	real	unrelated	NA	*third*	*contact*	NA	67037	5	25218	7	not controlled
L2_D	NS_2	real	unrelated	NA	*shove*	*human*	NA	6327	5	101224	5	not controlled
L2_D	NS_2	real	unrelated	NA	*deer*	*loud*	NA	11665	4	10324	4	not controlled
L2_D	NS_2	real	unrelated	NA	*mix*	*thank*	NA	20642	3	88574	5	not controlled
L2_D	NS_2	real	unrelated	NA	*warn*	*sauce*	NA	19996	4	15903	5	not controlled
L2_D	NS_2	real	unrelated	NA	*content*	*food*	NA	21821	7	107728	4	not controlled
L2_D	NS_2	pseudo	NA	NA	*legal*	*glur*	NA	44820	5	NA	4	aj
L2_D	NS_2	pseudo	NA	NA	*bitter*	*cinker*	NA	8353	6	NA	6	aj
L2_D	NS_2	pseudo	NA	NA	*cute*	*vone*	NA	6089	4	NA	4	aj
L2_D	NS_2	pseudo	NA	NA	*open*	*alk*	NA	111857	4	NA	3	aj
L2_D	NS_2	pseudo	NA	NA	*thin*	*prunch*	NA	23194	4	NA	6	aj
L2_D	NS_2	pseudo	NA	NA	*tribal*	*prue*	NA	6995	6	NA	4	aj

L2_D	NS_2	pseudo	NA	NA	*labor*	*stap*	NA	34400	5	NA	4	aj
L2_D	NS_2	pseudo	NA	NA	*sudden*	*baze*	NA	11370	6	NA	4	aj
L2_D	NS_2	pseudo	NA	NA	*national*	*feps*	NA	166359	8	NA	4	aj
L2_D	NS_2	pseudo	NA	NA	*coastal*	*slonk*	NA	6880	7	NA	5	aj
L2_D	NS_2	pseudo	NA	NA	*village*	*umity*	NA	30741	7	NA	5	n
L2_D	NS_2	pseudo	NA	NA	*root*	*fike*	NA	20774	4	NA	4	n
L2_D	NS_2	pseudo	NA	NA	*hole*	*theath*	NA	32302	4	NA	6	n
L2_D	NS_2	pseudo	NA	NA	*bank*	*rinch*	NA	58992	4	NA	5	n
L2_D	NS_2	pseudo	NA	NA	*climate*	*lape*	NA	17189	7	NA	4	n
L2_D	NS_2	pseudo	NA	NA	*crazy*	*pister*	NA	20345	5	NA	6	n
L2_D	NS_2	pseudo	NA	NA	*cup*	*piddy*	NA	53633	3	NA	5	n
L2_D	NS_2	pseudo	NA	NA	*bird*	*scell*	NA	34835	4	NA	5	n
L2_D	NS_2	pseudo	NA	NA	*soul*	*gushion*	NA	22713	4	NA	7	n
L2_D	NS_2	pseudo	NA	NA	*lawyer*	*tay*	NA	47853	6	NA	3	n
L2_D	NS_2	pseudo	NA	NA	*feather*	*mennel*	NA	5744	7	NA	6	n
L2_D	NS_2	pseudo	NA	NA	*mirror*	*umset*	NA	20081	6	NA	5	n
L2_D	NS_2	pseudo	NA	NA	*foot*	*maper*	NA	107285	4	NA	5	n
L2_D	NS_2	pseudo	NA	NA	*door*	*wun*	NA	124993	4	NA	3	n
L2_D	NS_2	pseudo	NA	NA	*moment*	*fam*	NA	109720	6	NA	3	n
L2_D	NS_2	pseudo	NA	NA	*yell*	*tashable*	NA	13131	4	NA	8	n
L2_D	NS_2	pseudo	NA	NA	*sister*	*derode*	NA	48183	6	NA	6	n
L2_D	NS_2	pseudo	NA	NA	*test*	*fays*	NA	69870	4	NA	4	n
L2_D	NS_2	pseudo	NA	NA	*gun*	*ploic*	NA	47305	3	NA	5	n
L2_D	NS_2	pseudo	NA	NA	*crime*	*goma*	NA	48010	5	NA	4	n
L2_D	NS_2	pseudo	NA	NA	*church*	*tefine*	NA	59466	6	NA	6	n
L2_D	NS_2	pseudo	NA	NA	*fool*	*trazy*	NA	6990	4	NA	5	n
L2_D	NS_2	pseudo	NA	NA	*nation*	*hemory*	NA	97212	6	NA	6	n
L2_D	NS_2	pseudo	NA	NA	*history*	*agongst*	NA	114904	7	NA	7	n
L2_D	NS_2	pseudo	NA	NA	*ceiling*	*gurbine*	NA	12668	7	NA	7	n
L2_D	NS_2	pseudo	NA	NA	*money*	*ralking*	NA	164794	5	NA	7	n
L2_D	NS_2	pseudo	NA	NA	*dust*	*chorty*	NA	15475	4	NA	6	n

L2_D	NS_2	pseudo	NA	NA	*twin*	*voon*	NA	14610	4	NA	4	n
L2_D	NS_2	pseudo	NA	NA	*diary*	*pefty*	NA	5945	5	NA	5	n
L2_D	NS_2	pseudo	NA	NA	*second*	*thark*	NA	103621	6	NA	5	n
L2_D	NS_2	pseudo	NA	NA	*bunch*	*unible*	NA	12225	5	NA	6	n
L2_D	NS_2	pseudo	NA	NA	*guest*	*hict*	NA	29328	5	NA	4	n
L2_D	NS_2	pseudo	NA	NA	*director*	*destive*	NA	79813	8	NA	7	n
L2_D	NS_2	pseudo	NA	NA	*fee*	*rin*	NA	20263	3	NA	3	n
L2_D	NS_2	pseudo	NA	NA	*distance*	*gairway*	NA	31380	8	NA	7	n
L2_D	NS_2	pseudo	NA	NA	*evening*	*yoor*	NA	40881	7	NA	4	n
L2_D	NS_2	pseudo	NA	NA	*snow*	*griby*	NA	21011	4	NA	5	n
L2_D	NS_2	pseudo	NA	NA	*jungle*	*mogus*	NA	5693	6	NA	5	n
L2_D	NS_2	pseudo	NA	NA	*priest*	*hedding*	NA	14889	6	NA	7	n
L2_D	NS_2	pseudo	NA	NA	*adult*	*cham*	NA	40705	5	NA	4	n
L2_D	NS_2	pseudo	NA	NA	*explore*	*harty*	NA	23744	7	NA	5	v
L2_D	NS_2	pseudo	NA	NA	*learn*	*mauper*	NA	124346	5	NA	6	v
L2_D	NS_2	pseudo	NA	NA	*sweep*	*prifle*	NA	12431	5	NA	6	v
L2_D	NS_2	pseudo	NA	NA	*grasp*	*vall*	NA	6357	5	NA	4	v
L2_D	NS_2	pseudo	NA	NA	*boast*	*gutor*	NA	5731	5	NA	5	v
L2_D	NS_2	pseudo	NA	NA	*remember*	*tampsite*	NA	106879	8	NA	8	v
L2_D	NS_2	pseudo	NA	NA	*ensure*	*massle*	NA	21341	6	NA	6	v
L2_D	NS_2	pseudo	NA	NA	*insist*	*woond*	NA	27205	6	NA	5	v
L2_D	NS_2	pseudo	NA	NA	*climb*	*hestow*	NA	24367	5	NA	6	v
L2_D	NS_2	pseudo	NA	NA	*prevent*	*mucrative*	NA	36421	7	NA	9	v

Appendix G: Video Stimuli for Study 3

Set		Title (adapted from original)	Edited Length	Original Source	Source URL
1	Practice	Octopus Love	01:20	Gobelins l'École de l'Image & Bocabeille et al. (2007)	https://www.youtube.com/watch?v=badHUNl2HXU
	Trial	Mouse for Sale	03:24	Bongaerts (2010)	https://www.youtube.com/watch?v=OzYwE3Tst1Y
2	Practice	Chess Player	01:52	Pixar & Pinkava, J. (1997)	https://www.youtube.com/watch?v=9IYRC7g2ICg
	Trial	Can I Stay	03:17	Lo, Carter, & Knudson (2015)	https://www.youtube.com/watch?v=im0k9d-gqbU
3	Practice	Tree House	01:58	Mugica, Hawkins, & Kang (2015)	https://www.youtube.com/watch?v=y47-gmGvZhI
	Trial	Carrot Crazy	03:15	Vanwormer & Scelina (2015)	https://www.youtube.com/watch?v=7V7MOk0FZrg
4	Practice	Jinky & Lucky	01:45	Bidinger & Kwon (2015)	https://www.youtube.com/watch?v=g4PnrN2EdAE
	Trial	Aviator	03:46	Yu & Tzue (2014)	https://www.youtube.com/watch?v=UUlaseGrkLc

Appendix H: Transcription Guidelines for Study 3

Your task is to transcribe the attached audio files. You will type (almost) everything you hear in (mostly) complete sentences.

All speakers are responding to one of 8 short videos (attached separately). Before beginning, you should watch each video at least once. Doing so will make it easier to understand what speakers are saying, especially if audio quality is low or if speakers are non-native speakers of English.

Please transcribe all **verbal** and **non-verbal fillers** because they may be of interest later on. You should type **verbal fillers** (e.g., *um*, *ah*, *oh*) exactly as they are spelled below without any brackets. Meanwhile, **non-verbal fillers** should be enclosed by square brackets.

See **punctuation** and **spelling** notes below for specific instructions on each.

See **Saving Transcriptions** at the bottom of this document for instructions on how to format and title your transcription files.

Meta-Data

- Please type speaker's participant ID number in square brackets: e.g., [1042]
- After enclosing the ID number by itself in square brackets, you should also type the first sentence or so in which the speaker may introduce him/herself and speak his/her ID number aloud. For example:
    - *[1042] [My participant ID number is 1042. My name is XX.] In this story, a man was playing chess in the park but he was alone.*
    - In the example above, you will notice that the speaker's name has been replaced with a double XX. Please do not type the speaker's name if it is included in the audio recording. Simply replace the name with a double XX.
    - To indicate if speaker was cut off at the end by the three minute time limit, simply type: [cut-off]

Verbal Fillers (do not use square brackets)

- *um* [ʌm]
- *uh* [ʌ]
- *ah* [ä]
- *eh* (probably rare); for clear [ɛ] rather than [ʌ] sound
- *er* (probably rare); for clear [ɚ] rather than [ʌ])
- *mmm* (thinking audibly)
- Separate *gonna* into *going to*
- Separate *kinda* into *kind of*
- Even if shortened to *cuz*, type *because*
- *oh* (as in *Oh no, wait…*)

Non-Verbal Fillers (please use square brackets)

- [???] to indicate inaudible speech (please type exactly three question marks)
- [cough]

- [sneeze]
- [clears throat]
- [laughs]

Punctuation

- Rely exclusively on commas, periods, question marks, and exclamation marks
- Avoid dashes or colons
- Exclamation marks should only be used for especially emphatic speech
- Use question marks only to indicate direct speech (*And then she was like, "What happened?")* or if the speaker asks him or herself a question aloud (e.g., *Oh, wait, what do you call that?*)
- When narrative is highly fluent, continuous speech, use a comma between independent clauses connected with *and*, *so*, or *but*. However, when intonation indicates a break and is accompanied by a perceptible pause, use a period and begin *And*, *So*, or *But* with a capital letter. The grammar rule to avoid beginning a sentence with a conjunction does not apply here.
- Use quotation marks when speaker is relating direct speech (*He was like, "Cut it out"*), even if you don't suspect that the speech actually occurred in the film.
- Only include whole words. So, if the speaker utters a partial word (e.g., *...she ke- keeps failing at it*) you should type, "…she keeps failing at it." However, if a speaker repeats an entire word in its entirety (e.g., *And then he he he he say…)* go ahead and type the word multiple times.

Spelling

- Refer to Merriam-Webster.com for spelling questions (e.g., whether a word should be hyphenated, whether a word)
- Spelling notes specific to videos

AVIATOR:
"airplane" (one word)
"airship" (one word)

PET SHOP:
"roly-poly" (hyphenate)
"teeter-totter" (hyphenate)
"pet shop" (two separate words)

HOMELESS:
"trash can" (two words)

OCTOPUS:
Do not correct "octopi" or "octopuses," just type whichever the speaker says

TREE HOUSE:
"tree house" (two words)

CHESS:

“Aha!" (if speaker reports man’s interjection)

Saving Transcriptions

Please save your transcription as a .txt file. If you don’t know how to do this, please ask. When you save the file, use the exact same file name as the audio. However, please underscore+your initials to the end of the file.

For example, if the audio file has the title 002_2017.02.17_octopus.mp3 and your initials are “CB” your transcription filename would be 002_2017.02.17_octopus_CB.txt

Questions?

- Ask!
- Cindy Berger
    - 470-230-7988
    - cberger@gsu.edu