THE POLYSEMY ADVANTAGE IN LEXICAL ACCESS: THE ROLE OF CONTEXT AVAILABILITY AND ORTHOGRAPHIC NEIGHBORHOOD VARIABLES

by

Caitlin A. Rice

B.A., Psychology, Grinnell College, 2011

Submitted to the Graduate Faculty of

The Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science in Cognitive Psychology

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Caitlin A. Rice

It was defended on

March 8, 2017

and approved by

Scott Fraundorf, Assistant Professor, Department of Psychology

Michael Dickey, Associate Professor, Department of Communication Science and Disorders

Thesis Director: Natasha Tokowicz, Associate Professor, Departments of Psychology and Linguistics

The polysemy advantage in lexical access: The role of context availability and orthographic neighborhood variables

Caitlin A. Rice, M.S.

University of Pittsburgh, 2017

In this study we examine the interactions between lexical and semantic variables during lexical decision. We replicate findings that semantically ambiguous words with related senses (polysemes) have an advantage in lexical decision. However, we report a reversal of the traditional concreteness effect, such that there is a disadvantage for words high in concreteness. Furthermore, we report an advantage for words high in context availability but note that this advantage is qualified by both contextual diversity and orthographic neighborhood frequency. In contrast to past findings, ambiguity and context availability did not interact, although a novel interaction of ambiguity and contextual diversity is reported. We discuss the implications of these findings for the context availability hypothesis (Schwanenflugel & Shoben, 1983).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

The helpful comments of Natasha Tokowicz, PhD, Scott Fraundorf, PhD, Mike Dickey, PhD, And Blair Armstrong, PhD are gratefully acknowledged. In addition, we thank Teljer L. Liburd for study design assistance, and Chad R. Tokowicz and the PLUM Lab research assistants for assistance with data collection and coding.

1.0 INTRODUCTION

A majority of words in the English language have multiple meanings depending on the contexts in which they occur (Klein & Murphy, 2001). Such *semantically ambiguous* words have received a great deal of attention over the past few decades (for a review, see Eddington & Tokowicz, 2015), and it is now well-established that ambiguous words are recognized more quickly and accurately in lexical decision tasks than unambiguous words (e.g., Azuma & van Orden, 1997; Hino, Pexman, & Lupker, 2006; Jaztrzembski, 1981). However, it was recently discovered that this ambiguity advantage depends on the relatedness of a word's meanings (Rodd, Gaskell, & Marslen-Wilson, 2002). Ambiguous words can be divided based on meaning relatedness into two main types: *homonyms* and *polysemes*. Homonyms are ambiguous words with unrelated meanings (e.g., LOCK can mean either *a device to prevent entry* or *a curl of hair*), whereas polysemes are ambiguous words whose meanings are semantically related (e.g., FOOT can mean either *a body part* or *a unit of measure*). It was originally thought that homonyms had an advantage relative to unambiguous words in word recognition as revealed by lexical decision tasks, but recent research indicates that in fact they have a disadvantage (e.g., Armstrong & Plaut, 2008, 2011; Beretta, Fiorentino, & Poeppel, 2005) or no advantage (Rodd et al., 2002; Experiment 2) in lexical decision. In contrast to homonyms, polysemes have an advantage in lexical decision (e.g., Jager & Cleland, 2014; Klepousniotou, 2002; Rodd et al., 2002), although this advantage is sensitive to the effects of lexical and semantic variables (Azuma & van Orden, 1997; Jager & Cleland, 2014; Tokowicz & Kroll, 2007). The current study aims to: (a) examine the influence of ambiguity and lexical and semantic variables of theoretical interest, namely concreteness, contextual diversity, and orthographic neighborhood features, on word recognition during lexical decision, (b) provide an inclusive account of the complex interactions between lexical and semantic variables using linear mixed effects modeling, and (c) discuss the implications of our findings for theoretical accounts of semantic ambiguity. We first review models of semantic ambiguity in lexical processing to lay the theoretical groundwork for this study. We then review the effects and interactions of four variables known to influence lexical processing—semantic ambiguity, word concreteness, orthographic neighborhood, and contextual diversity. Finally, we discuss the implications of our findings for models of semantic ambiguity and concreteness effects in lexical processing.

1.1 MODELS OF SEMANTIC AMBIGUITY IN LEXICAL PROCESSING

The mixed pattern of results produced by studies of lexical processing has been addressed by models of semantic ambiguity that explain the source of the polysemy advantage with regard to semantic activation and settling dynamics during the time course of lexical processing. For instance, Armstrong and Plaut (2016) proposed the Semantic Settling Dynamics (SSD) account of semantic ambiguity resolution within a biologically-plausible connectionist model. The SSD proposes that there are differences in the neural representations of polysemous, homonymous, and unambiguous words, and that these kinds of words are processed in distinct and predictable ways that shift over the time course of lexical processing. The semantic features of polysemes overlap to a large extent whereas the semantic features of homonyms overlap very little. Nonoverlapping features compete in the word recognition process, and competition must be resolved for a word to reach a recognition threshold. The SSD proposes that polysemes will benefit from early excitatory feedback arising from cooperation between two or more semantically-related senses, whereas homonyms will initially be hindered by competing semantically-unrelated meanings, and must wait for inhibitory connections to come online later in processing before a decision can be made between two or more competing meanings.

The predictions of the SSD have been simulated via a connectionist model with input patterns representing unambiguous words, homonyms, and polysemes, and tested against human lexical decision latencies (Armstrong & Plaut, 2008). After training, the model produced an early processing advantage for polysemous words over homonyms and unambiguous words, which is generally consistent with behavioral results reported in earlier research (e.g. Armstrong & Plaut, 2011).

1.2 CONCRETENESS IN LEXICAL PROCESSING

We now turn to discussing semantic variables that may interact with ambiguity during lexical processing. One commonly investigated semantic effect is the *concreteness effect*; that is, the fact that concrete words are recognized more quickly than abstract words. Although the concreteness effect is often thought of as a ubiquitous effect in lexical decision (James, 1975; Schwaneflugel, Harnishfeger, & Stowe, 1988), free recall (Paivio, 1971), sentence processing (Begg & Paivio, 1969), and bilingual processing (van Hell & de Groot, 1998a, 1998b), the literature in this area describes a complex and sometimes contradictory pattern of results, including numerous studies that have failed to find a concreteness effect (e.g., Kroll & Merves, 1986; Samson & Pillon, 2004; Schwaneflugel et al., 1988; Tokowicz & Kroll, 2007; Tolentino & Tokowicz, 2009) and a handful that have reported a reversal of the effect (e.g., Tokowicz & Kroll, 2007; Vigliocco et al., 2014).

One theory that may explain inconsistent responses to concrete words in lexical decision is the context availability hypothesis (Schwanenflugel & Shoben, 1983), which attributes the concreteness effect to the ease of thinking of a context in which a word could occur. According to this hypothesis, it is easier to retrieve such contextual information for concrete words than for abstract words on average, and as a result concrete words tend to be processed more quickly than abstract words. This model predicts that when context availability is controlled, the concreteness advantage will be eliminated. Indeed, Schwanenflugel and Shoben (1983) demonstrated an advantage in lexical decision times for concrete versus abstract nouns that disappeared when these abstract and concrete words were embedded in supportive context sentences. Similarly, Schwaneflugel et al. (1988) found that the concreteness effect disappeared when context availability was controlled, and that context availability was a better predictor of lexical decision latencies than concreteness, imageability, familiarity, or age of acquisition.

A second theory that addresses concreteness effects is the dual-coding theory (Paivio, 1971, 1986). According to this theory, words are represented in memory both visually (imagistically) and verbally, and are processed by distinct and specialized subsystems. Concrete words have both imaginal and verbal representations, whereas abstract words have primarily verbal representations with few or no imaginal representations. Thus, concrete words have an advantage because they activate two sources of information and two processing systems, whereas abstract words primarily rely on one representation and one processing system. This model predicts that concreteness effects would not be present for words with equal amount of imaginal information. This theory was originally developed to describe imagability effects in memory tasks, and it has been noted that imaginal information may not come online early enough to influence processing in relatively rapid tasks such as lexical decision (e.g., Tolentino & Tokowicz, 2009). However, Schwanenflugel and Stowe (1989) extended the dual-coding theory to word recognition tasks such as lexical decision and referred to this version as the dual-representation view. Therefore, we test this version of the theory, which we refer to as the DRV, in the present study.

1.3 INVESTIGATING LEXICAL AND SEMANTIC VARIABLES

The inclusion of both concreteness and context availability in a single model will allow us to test the predictions of both the context availability hypothesis and the DRV. The context availability hypothesis predicts that when context availability is controlled, concreteness effects should be eliminated. Conversely, the DRV predicts that concreteness effects arise from the greater imageability of concrete vs. abstract words, and so if this hypothesis is correct then concreteness effects will remain even when context availability is controlled.

However, investigating the contributions of these and other lexical and semantic variables known to influence word recognition can be difficult given that a large number of highly correlated variables are of interest, and multicollinearity is a concern in studies that have used multiple regression to investigate similar variables and effects. For instance, the high intercorrelations[1] of concreteness, imageability, and context availability have been a source of confusion in the literature, and researchers have handled this in different ways. In some cases, researchers have discussed imageability and concreteness interchangeably (Paivio, 1968; Reilly

& Kean, 2007; van Hell & de Groot, 1998b). In other cases, researchers have preferred concreteness over imageability because concreteness tends to have 1) a more binomial distribution, which makes it a better predictor for categorical variables, and 2) greater variance which makes for a better linear predictor, because restricted ranges may attenuate correlations or lead to floor and ceiling effects (Paivio, 1971; Kousta et al., 2011).

Similarly, concreteness and context availability tend to be correlated. Here, we attempt to disentangle some of their effects by including both concreteness and context availability in our linear mixed effects models. We noted that in our sample concreteness and context availability are highly correlated (r = .76; see Table 3), so we examined the condition number, a test of the overall amount of collinearity in a model (Baayen, 2008). The condition number indicated that there was medium collinearity (κ = 20.59), but this value falls below the threshold for potentially harmful collinearity (κ = 30; Baayen, 2008). We then examined the variance inflation factor (VIF), a measure of how much individual variables in a model are affected by collinearity (Frank, 2011). A VIF value over 5.00 is potentially problematic, and we observed a value of 6.12 for context availability and 5.56 for concreteness. To determine if these variables were redundant in our model, we conducted tests of model fit and found that a model with concreteness and context availability fit better than a model with only concreteness or a model with only context availability.[2] Therefore, while we acknowledge that multicollinearity needs to be carefully examined in studies of the lexical and semantic variables that affect lexical decision, we believe the inclusion of both context availability and concreteness in out model provides valuable information and the independent effects of each variable can still be measured.

In addition to the correlations among key variables, a second challenge for studying ambiguity and concreteness effects is that lexical and semantic variables may interact during lexical processing, and past studies have often not examined these interactions. For example, polysemy and concreteness effects have been studied extensively, but few studies have examined if these independently strong predictors of lexical decision latencies interact during lexical processing. The first study to examine this possibility was conducted by Tokowicz and Kroll (2007), who tested whether concreteness/context availability and ambiguity interact during lexical decision. In their Experiment 3, native English speakers completed a lexical decision task with words that varied in concreteness, context availability, and number of meanings. Tokowicz and Kroll reported an interaction of ambiguity and context availability such that the effects of ambiguity were only present for words low in context availability. However, the words in this study included homonyms as well as polysemes (for additional examples of cases in which ambiguous words were likely polysemous words, see Azuma & van Orden, 1997; Millis & Button, 1989), which complicates interpretation of the results.

Additionally, the study used Kucera and Francis (1967) word frequency norms, which Brysbaert and New (2009) argued are a weaker predictor of lexical decision latencies than other sources now available. Finally, the study used hierarchical linear regression models accounting for variance only across items but not across subjects. The present study aims to improve on these issues and test whether an interaction between ambiguity and concreteness/context availability is present after excluding homonyms from the word set, using contextual diversity as an updated measure of word frequency (Adelman, Brown, & Quesada, 2006) and analyzing the results using linear mixed effects models to allow us to account for both subject and item variance (Baayen, Davidson, & Bates, 2008).

The need to test for interactions between ambiguity and concreteness/context availability has been addressed to some degree by Jager and Cleland (2014), who conducted two lexical decision test experiments in a word set comprising only polysemes. In Experiment 1, they were unable to replicate the polysemy advantage in a set of concrete nouns, half of which had few (1-4) senses and half of which had many (6 or more) senses. In Experiment 2, they developed a second word set that was matched on lexical features to set 1, but that instead of concrete nouns comprised a mix of abstract nouns, verbs, adjectives, and adverbs. Half of these words had few senses and half had many senses, and in this word set they reported a significant polysemy advantage. The authors presented these results as replicating the ambiguity and concreteness/context availability interaction of Tokowicz and Kroll (2007), because the polysemy advantage was present for concrete but not abstract nouns. However, Jager and Cleland (2014) had several study design issues that the current study will seek to address. First, they used a non-continuous measure of ambiguity, which many researchers consider undesirable because dichotomizing a continuous variable reduces the amount of information available and may result in spurious results or a failure to find significant effects (for a review see: Tokowicz & Warren, 2008). Second, the study did not directly test for an interaction between polysemy and concreteness, but rather demonstrated a polysemy advantage for abstract words but not concrete nouns in separate stimulus sets. It is possible that these stimulus sets varied in some unanticipated way for which they were not matched. For example, although stimuli were matched for number of orthographic neighbors, they were not matched for orthographic neighborhood frequency, which has direct effects on lexical decision latencies and interacts with concreteness (Samson & Pillon, 2004). In addition, stimuli were not matched for context availability, which is problematic because previous research has demonstrated that including context availability in a model can eliminate concreteness effects (Tokowicz & Kroll, 2007). Third, the concrete words were only nouns whereas the abstract words included nouns, verbs, adjectives, and adverbs; this is problematic because grammatical class can influence ambiguity effects (Armstrong & Plaut, 2016; Mirman, Strauss, Dixon, & Magnuson, 2010).

1.4 EFFECTS OF ORTHOGRAPHIC NEIGHBORHOOD

Theories of ambiguity and concreteness have largely focused on differences in the semantic structure of concrete vs. abstract words, but orthographic neighborhood characteristics may help explain semantic ambiguity and concreteness effects. *Orthographic neighborhood,* as defined by Coltheart's *N* metric (Coltheart, 1977), is the number of real words that can be derived from a target word by substituting one letter, while holding the number and position of the other letters constant. Orthographic neighborhood density is the number of neighbors a target word has, and orthographic neighborhood frequency is the number of higher frequency neighbors a target word has. Samson and Pillon (2004) found evidence that words with higher orthographic neighborhood density are recognized more quickly than words with lower orthographic neighborhood density, whereas words with higher orthographic neighborhood frequency are recognized more slowly than words with lower orthographic neighborhood frequency. Taken together, the facilitative effect of orthographic neighborhood density and the inhibitory effect of orthographic neighborhood frequency in lexical decision indicate that orthographic neighborhood characteristics must be controlled in order to study semantic effects in lexical decision.

In addition to main effects of orthographic neighborhood variables, Samson and Pillon (2004) reported an interaction of orthographic neighborhood frequency (but not orthographic neighborhood density) with concreteness, such that the typical concreteness effect emerged for words with one or more higher frequency orthographic neighbors, but no concreteness effect was found for words with no higher frequency orthographic neighbors.

Although the interaction of orthographic neighborhood frequency and concreteness has only been reported by Samson and Pillon (2004), other researchers have reported that orthographic neighborhood interacts with both word frequency and ambiguity. For instance, Andrews (1989) found evidence that in lexical decision tasks, low frequency words are responded to more quickly the larger their orthographic neighborhood, whereas higher frequency words are less influenced by neighborhood size. Further, Ferraro and Hansen (2002) reported a three-way interaction of orthographic neighborhood density, orthographic neighborhood frequency, and ambiguity, such that the ambiguity effect was strongest for words with many orthographic neighbors, but only when these words had many higher frequency neighbors in their neighborhoods. The work of Samson and Pillon (2004), Andrews (1989; 1992), and Ferraro and Hansen (2002) has highlighted the possibility of orthographic neighborhood variables interacting with both semantic ambiguity and concreteness, and so the current study will test for these interactions.

1.5 LEXICAL FREQUENCY

In addition to orthographic neighborhood variables, lexical frequency affects word recognition times, and may also interact with key variables in this study. It is a reliable finding that higher frequency words are recognized more quickly than lower frequency words (Broadbent, 1967; Forster & Chambers, 1973). Traditionally, the word frequency effect is thought to be due to the fact that each encounter with a word strengthens memory traces and makes words easier to access in the future. Higher frequency words are encountered more often than low frequency words, and so are accessed more quickly (i.e., the repetition effect). Word frequency occupies a central place in many models of word recognition, such as the dual-route model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), the Interactive Activation Model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), and the EZ Reader Model (Reichle, Pollatsek, Fisher, & Rayner, 1998).

There are also several reports of word frequency interacting with ambiguity. In one of the earliest reports of this interaction, Rubenstein, Garfield, and Millikan (1970) conducted a lexical decision task with homographs and nonhomographs that varied in word frequency, and reported a significant interaction of frequency and ambiguity, such that the effect of ambiguity was significantly greater for low-frequency than high-frequency homographs. Jaztrzembski (1981) expressed concern that Rubenstein et al.'s homographs were not sufficiently different from their nonhomographs, and conducted a follow-up study that instead used words with high and low numbers of meanings and high and low frequencies in a lexical decision task. Jaztrzembski also reported a significant interaction between ambiguity and frequency in which the effects of ambiguity were more pronounced for low frequency than high frequency words, but interestingly, low frequency words with more meanings were recognized significantly more quickly than high frequency words with fewer meanings. Finally Pexman, Hino, and Lupker (2004; Experiment 1) conducted a visual lexical decision task with high- and low-frequency homonyms and unambiguous words, and reported an ambiguity advantage only for low-frequency words, which parallels the findings of Rubenstein et al. (1970).

Although word frequency is often thought of as the strongest predictor of word recognition, the dominance of this effect has recently been challenged. Baayen (2010) argued that although word frequency has commonly been reported as the strongest predictor of variation in visual lexical decision performance, 90% of this variation is in fact predicted by other lexical characteristics. Baayen conducted a principal components analysis with 17 lexical factors (e.g., length, bigram frequency, contextual diversity, neighborhood density, etc.) known to affect lexical processing, and the component that explained the most variance (28.1%) in lexical decision included several measures of the diversity of the contexts in which words appear. This component outperformed the component onto which word frequency loaded, leading Baayen (2008) to conclude that the importance of simple frequency of occurrence has been overestimated, whereas the importance of diversity of occurrences has been underestimated.

Several additional studies support the idea that the simple number of encounters with a word is not as critical for word recognition as the number of different contexts in which a word appears (Adelman et al., 2006; Baayen, 2010; Perea, Soares, & Comesana, 2013). Adelman et al. (2006) developed a measure called contextual diversity, which they defined as the number of corpus passages or documents that contain a target word. They reported that in regression models predicting lexical decision latencies, contextual diversity captured the variance associated with word frequency and also uniquely predicted additional variance. On this basis, they concluded that word frequency effects are artifactual, and in actuality contextual diversity is more predictive. Perea et al. (2013) conceptually replicated this finding in a sample of Portuguese-speaking children. However, these studies did not test whether concreteness/context availability interacts with contextual diversity, which one might expect given that word frequency and concreteness are known to interact (de Groot, 1989; James, 1975). To the best of our knowledge this hypothesis has not yet been tested, and thus is a novel contribution of the present work.

1.6 PRESENT STUDY

In the current study, we examine the pattern of interactions between lexical variables (orthographic neighborhood frequency, orthographic neighborhood density) and semantic variables (concreteness, context availability, polysemy, contextual diversity) in a visual lexical decision task. In contrast to past examinations of ambiguity and concreteness effects, we carefully selected non-homonymous polysemes (i.e., words that have more than one sense but only one meaning) that vary in number of senses (NOS).

The first aim of this study is to test whether there is an interaction between polysemy (hereafter NOS) and concreteness/context availability. We hypothesize that the polysemy effect is qualified by concreteness/context availability such that there is only an advantage for polysemous words that are low in concreteness/context availability. If this is indeed what we find, it will provide support for Armstrong and Plaut (2016)'s SSD account, which predicts initial excitation from the non-competing senses of polysemous words speeds recognition early in lexical processing. We predict that words that are low in concreteness/context availability (i.e., words that have fewer sensory referents and semantic features) are particularly difficult to recognize, and so may get an extra boost from being high in NOS. We further predict that words that are low in NOS and also low in concreteness/context availability will be recognized most slowly.

The second aim of this study is to test the context availability hypothesis, which predicts that we will find significantly faster lexical decision time for words that are higher in context availability, even after controlling for concreteness and other variables known to explain variation in lexical decision latencies. Additionally, it predicts that concreteness effects will disappear when context availability is controlled. If we do not find an advantage for words high in context availability this contradict the predictions of the context availability hypothesis and may offer support for DCT-LD, which predicts that concreteness effects are due to differential levels of imagistic information between abstract and concrete words, and so the concreteness effect should persist when context availability is controlled.

The third aim of the study is to determine the lexical variables to which polysemy and concreteness/context availability effects are sensitive. Based on findings from Samson and Pillon (2004), we expect the orthographic neighborhood variables may qualify these effects and their hypothesized interaction. Specifically, we expect to find an interaction between orthographic neighborhood frequency and polysemy, such that words with few senses will be particularly vulnerable to inhibition from higher frequency orthographic neighbors. Additionally, based on the results of Adelman et al. (2006) and previous reports of an interaction between concreteness and word frequency (de Groot, 1989; James, 1975) we hypothesize that the concreteness/context availability effect may be qualified by contextual diversity, such that the concreteness/context availability effect is most apparent for words low in contextual diversity.

2.0 METHOD

2.1 PARTICIPANTS

Participants were 102 monolingual English speakers recruited from the University of Pittsburgh Department of Psychology participant pool, who were given credit towards a class requirement for their participation. Data from 20 participants were excluded: 17 due to low accuracy, two due to equipment malfunctions during testing, and one due to missing data. Data from 82 participants remained for analysis. All participants were 18 years of age or older, right-handed, and had normal or corrected-to-normal vision.

2.2 STIMULI

Stimuli were a subset of 497 words from the stimulus lists used by Tokowicz and Kroll (2007) and 497 pseudowords matched to the words on word length, number of orthographic neighbors, and bigram frequency (see Table 1). Forty-five of the original 497 words (9.1% of items) were removed from analyses because data extracted from WordNet (Miller, 1995) revealed that they had more than one meaning (i.e., were homonyms). The remaining 452 words had exactly one meaning, but varied in NOS.

Table 1. Stimulus properties

	Words		Pseudowords	
	Mean	SD	Mean	SD
Length (number of letters)	5.87	1.67	5.79	1.68
Number of senses (NOS)	5.26	3.91	-	-
Orthographic neighborhood frequency	0.68	1.36	-	-
Orthographic neighborhood density	3.10	4.18	3.27	4.42
Summated bigram frequency	1.62	0.89	1.57	0.86
Concreteness	4.56	1.76	-	-
Context availability	5.85	0.61	-	-
Contextual diversity	17.41	19.97	-	-

NOS data were extracted from WordNet, and ranged from 1 to 35 (M = 5.26, SD = 3.91). To check the reliability of these data, NOS data from WordNet were compared to NOS from the Wordsmyth Online Dictionary (collected by the eDom project; Armstrong, Tokowicz, & Plaut, 2012), and the two were highly correlated (r = .98) for this sample. An item-level comparison revealed that discrepancies in NOS between the two databases were no greater than two senses, except for one word with a discrepancy of 6 senses (*respect*). This item was removed, leaving a total of 451 words for the analyses.

Pseudowords were constructed by querying the eLexicon database (Balota et al., 2007) for nonwords that were matched to the Tokowicz and Kroll (2007) real words on length, orthographic neighborhood density, and bigram frequency. This method yielded English-like pseudowords (i.e., pseudowords that were pronounceable and had no illegal bigrams).

Normative ratings were collected for concreteness and context availability of the real words from undergraduate students at the University of Pittsburgh; one group (N= 98) rated context availability and another (N = 98) rated concreteness. None of these participants had completed the primary task. Instructions for context availability ratings were adapted from Schwanenflugel and Stowe (1988). Participants were instructed to rate on a seven-point Likert-type scale how easy (1) or hard (7) it was for them to think of a context for a given word. Instructions for concreteness ratings were adapted from Spreen and Shulz (1966). Participants were instructed to rate words on a seven-point Likert-type scale from 1 (least concrete) to 7 (most concrete). Average concreteness and context availability ratings were calculated for each real word (see Appendix). Bigram frequency, orthographic neighborhood density, orthographic neighborhood frequency, and contextual diversity were extracted from the eLexicon project (Balota et al., 2007; see Table 1). For example stimuli, see Table 2.

Table 2. Example stimuli

	Low NOS	High NOS	Low CA	High CA
Low CD	stapler, tranquility, tuba, mittens	thread, feather, comb, flame	earnestness, allegory, malice, equity	panther, zebra, volcano, mustache
High CD	fun, clothes, hotel, kitchen	square, work, points, head	meaning, thing, stuff, situation	idea, room, house, money
Low ONF	-	-	temerity, elm, vanity, deceit	grandmother, fruit, computer, policeman
High ONF	-	-	folly, fate, ease	hat, rug, cart, meat

Abbreviations. *CA = context availability, CD = contextual diversity, NOS = number of senses, ONF = orthographic neighborhood frequency*

2.3 DESIGN AND PROCEDURE

Participants who provided concreteness and context availability ratings were instructed to rate all items using the full range of a 7-point Likert-type scale. Responses from participants who did not use the full scale (i.e., at least one “1” and one “7”) were excluded. Of the 98 respondents who rated concreteness, 27 were excluded (15 for missing data, nine for language background other than native English speaker, and three who did not use the full rating scale). Of the 98 respondents who rated context availability, 28 were excluded (four for having already responded to this survey, four for language background other than native English speaker, 16 for not using the full rating scale, and four for missing data). Thus, concreteness ratings were computed from a final set of 71 participants and context availability ratings were computed from a final set of 70 participants.

Participants completed a computerized visual lexical decision task in which they were presented with a word or pseudoword and responded by pressing a key with their dominant (right) hand for ‘word’ or another key with their non-dominant (left) hand for ‘nonword’. Each trial began with a fixation cross with a duration that varied randomly between 170 ms and 230 ms. Next a target word appeared on the screen until the participant responded or 3000 ms elapsed, whichever came first. All participants viewed all 497 words and 497 pseudowords, which were presented in randomized blocks of 100 trials with a self-timed break in between each block. All trials were presented on a PC running E-Prime Version 2.0 (Psychology Software Tools, 2012), and response time (in ms) and accuracy were collected. Following the lexical decision task, participants completed a language history questionnaire (Tokowicz, Michael, & Kroll, 2004) to confirm language background.

3.0 RESULTS

3.1 ANALYSIS PLAN

Reaction time data were analyzed using linear mixed effects models (Baayen, 2008) to account for random effects of both subjects and items, while examining experimental fixed effects that were of theoretical interest in the study. Ninety-five percent confidence intervals are reported for all fixed effects, and a conservative critical value of t = 2.0 was used for significance (Baayen et al., 2008). All analyses were conducted in R using version 1.1-7 of the lme4 package (Bates, Maechler, Bolker, & Walker, 2014).

3.2 DATA PROCESSING

Pseudowords were removed before analysis, and 1302 incorrect responses to real words were excluded (3.5% of trials). Data were excluded from participants with less than 90% accuracy on lexical decisions for word and pseudoword trials combined (e.g., Tokowicz & Kroll, 2007), which resulted in the removal of 17 participants (16.67% of the original sample). Prior to excluding these 17 participants, we examined accuracy for words vs. pseudowords separately, and found that pseudoword accuracy was below 90% for all 17 excluded participants whereas word accuracy was below 90% for only two of these 17 participants (87 and 89%). Responses faster than 200 ms were considered spurious and were therefore excluded, which resulted in the removal of 15 additional trials (.04% of correct word trials). Trials with response times 2.5 standard deviations or greater above or below a participant's mean response time for real words were excluded as outliers and treated as missing values, which resulted in the removal of 1119 trials (3.1% of correct word trials). Aside from these exclusions, there were no other missing data.

After exclusions, participants' mean accuracy ranged from 93% to 99% (M = 97%); that is, all participants answered nearly all trials correctly. Because of this near-ceiling performance in accuracy, we focus our analysis on the responses times for correct trials.

In all analyses, predictor variables of previous word response time, word length, bigram frequency, concreteness, context availability, contextual diversity, and NOS were mean-centered to eliminate non-essential multicollinearity (Jaeger, 2010). Bigram frequency and previous response time were scaled to match other variables by dividing by 1000, and contextual diversity was log-transformed.

3.3 WORD STATUS EFFECTS

Data for words and nonwords were compared using t-tests to examine word status effects on response time and accuracy. Consistent with past reports, words were responded to significantly more quickly (585.02 vs. 682.55 ms), $t_1(81) = 14.74$, $p < .001$; $t_2(946) = 28.71$, $p < .001$ and more accurately (96.63 vs. 93.13 % correct), $t_1(82) = 6.32$, $p < .001$; $t(946) = 6.70$, $p < .001$ than nonwords.

3.4 REAL WORD TRIALS

After verifying the presence of the expected real word advantage, we removed nonword trials and constructed a linear mixed effects model to investigate whether there is an ambiguity advantage for words with a greater NOS, and, if so, whether NOS interacts with lexical and semantic factors such as concreteness, context availability, contextual diversity, and orthographic neighborhood features. The dependent variable in this model was lexical decision latency (ms), and the following fixed effects and their interactions (up to 2-way interactions; there were no 3-way or higher interactions) were of theoretical interest: context availability, concreteness, contextual diversity, and NOS. Fixed effects of previous trial reaction time, previous trial accuracy, word length, orthographic neighborhood density, orthographic neighborhood frequency, and bigram frequency were included to control for variables known to be associated with word recognition. Random intercepts for subjects and items were included, and the fit of the model was improved significantly by including random by-subject slopes for context availability and contextual diversity, $\chi^2(2) = 127.47$, $p < .001$. The inclusion of random by-subject slopes for concreteness or NOS did not improve the model fit significantly, and therefore were omitted from the final model.[3]

The model equation and results for all variables entered in the model are presented in Table 3, and Table 4 shows the correlations among the fixed effects. Concreteness was a significant predictor of RT ($b = 6.04$, $SE = 1.95$, $t = 3.09$); there was a concreteness disadvantage in which every one point increase in concreteness results in a 6 ms (95% CI: 2 – 10) increase in RT. Although the direction of this effect is in contrast with the traditional concreteness advantage, it is consistent with findings reported by Tokowicz and Kroll (2007; Experiment 3). Concreteness did not interact significantly with any other variable of theoretical interest.

Table 3. Correlations of variables of theoretical interest

	ONF	OND	Concreteness	CA	NOS
OND	.71**				
Concreteness	.17**	.18**			
CA	.13**	.24**	.76**		
NOS	.03	.27**	.01	.12**	
CD	.02	.34**	.03	.37**	.42**

Abbreviations. *CA = context availability, CD = contextual diversity, NOS = number of senses, OND = orthographic neighborhood density, ONF = orthographic neighborhood frequency*
*** = correlation is significant at the .01 level*

Table 4. Fixed effects estimates

			95% CI			
	Estimate	SE	Lower Bound	Upper Bound	t	Significance
Intercept	641.75	7.95	626.17	657.33	80.72	*
Previous trial RT	79.51	2.85	73.92	85.09	27.9	*
Previous trial accuracy	-74.17	3.57	-81.17	-67.16	-20.76	*
Length	2.05	1.32	-0.54	4.65	1.55	
Bigram frequency	1.34	1.91	-2.39	5.07	0.7	
OND	-0.57	0.65	-1.84	0.7	-0.88	
CA	-38.64	6.4	-51.18	-26.09	-6.04	*
ONF	3.22	1.56	0.16	6.28	2.06	*
Concreteness	6.04	1.95	2.21	9.86	3.09	*
NOS	-1.36	0.44	-2.23	-0.49	-3.07	*
CD	-36.04	3.7	-43.29	-28.79	-9.74	*
OND * CA	4.26	2.17	0	8.51	1.96	
CA * ONF	-10.62	4.73	-19.89	-1.35	-2.25	*
OND * concreteness	-1.04	0.57	-2.16	0.07	-1.84	
ONF * concreteness	2.39	1.38	-0.32	5.1	1.73	
CA * NOS	-2.14	1.6	-5.28	1	-1.33	
Concreteness * NOS	0.83	0.51	-0.16	1.83	1.64	
NOS * CD	2.63	0.71	1.23	4.02	3.7	*
CA * CD	18.75	5.89	7.2	30.3	3.18	*
Concreteness * CD	-4.58	2.59	-9.65	0.49	-1.77	

Model equation. *Model1 <-lmer(RT ~ 1 + prevRT + prevACC + Length + BGFreq + OND*CA + ONF*CA+ OrthoNeighborhoodDensity*Concreteness + ONF*Concreteness + CA*NOS+ Concreteness*NOS + CD*NOS + CD*CA + CD*Concreteness + (1|Subject) + (1|Item) + (0 + CD + CA ||Subject))*
Abbreviations. *CA = context availability, CD = contextual diversity, OND = orthographic neighborhood density, ONF = orthographic neighborhood frequency*

In addition to a main effect of concreteness, there were also significant main effects of NOS, contextual diversity, context availability, and orthographic neighborhood frequency, all of which were in the expected directions. Words with a greater number of senses were recognized more quickly (b = -1.36, SE = 0.44, t = -3.07), as were words higher in contextual diversity (b = -36.04, SE = 3.7, t = -9.74), and higher in context availability (b = -38.64, SE = 6.4, t = -6.04). Words with higher frequency orthographic neighbors were slower to be recognized (b = 3.22, SE = 1.56, t = 2.06).

The main effects of context availability, NOS, contextual diversity, and orthographic neighborhood frequency were qualified by significant interactions. All significant interactions were probed by using the regression equation from Model 1 to generate estimated lexical decision latencies at one standard deviation above and below the mean for both fixed effects terms in the interaction (Aiken & West, 1991). First, there was a significant interaction of NOS and contextual diversity (b = 2.63, SE = 0.71, t = 3.7). For words low in NOS, there was a large disadvantage to being low in contextual diversity (M_{highCD} = 603.78 vs. M_{lowCD} = 655.65), whereas for words high in NOS the disadvantage to being low in contextual diversity was less pronounced (M_{highCD} = 604.66 vs. M_{lowCD} = 633.51; see Figure 1).

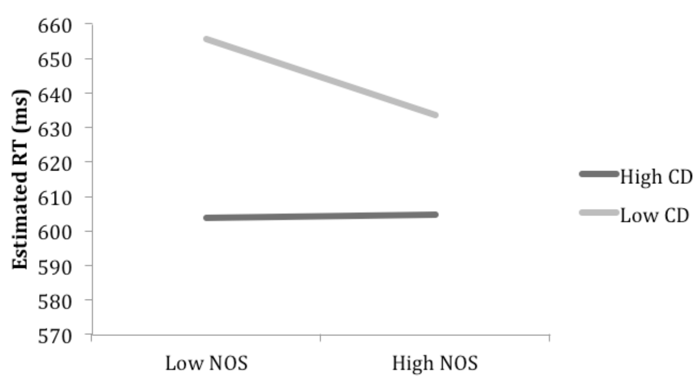

Figure 1. Estimated lexical decision latencies (ms) for high and low contextual diversity (CD) and number of senses (NOS)

Second, there was a significant interaction of context availability and contextual diversity (b = 18.75, SE = 5.89, t = 3.18; see Figure 2). For words high in context availability there was little effect of contextual diversity (M_{highCD} = 523.09 vs. M_{lowCD} = 508.79), whereas for words low in context availability, low contextual diversity words were responded to more slowly than high contextual diversity words (M_{highCD} = 517.57 vs. M_{lowCD} = 580.02).

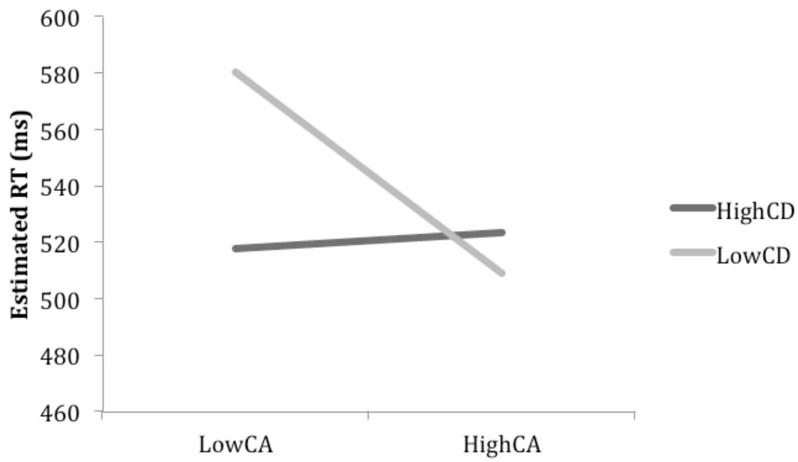

Figure 2. Estimated lexical decision latencies (ms) for high and low context availability (CA) and high and low contextual diversity (CD)

Finally, there was a significant interaction of context availability and orthographic neighborhood frequency (b = -10.62 , SE = 4.73, t = -2.25; see Figure 3). This interaction was not visually apparent when graphing the mean values +/- 1 SD, so instead we used the minimum and maximum values of context availability and orthographic neighborhood frequency observed in the sample. For words high in context availability there was little effect of orthographic neighborhood frequency (M_{highONF} = 491.49 vs. M_{lowONF} = 484.82), whereas for words low in context availability there was a disadvantage for words that had more higher frequency orthographic neighbors (M_{highONF} = 1140.63 vs. M_{lowONF} = 705.37). See Table 5 for the estimates of the random effects for Model 1.[4]

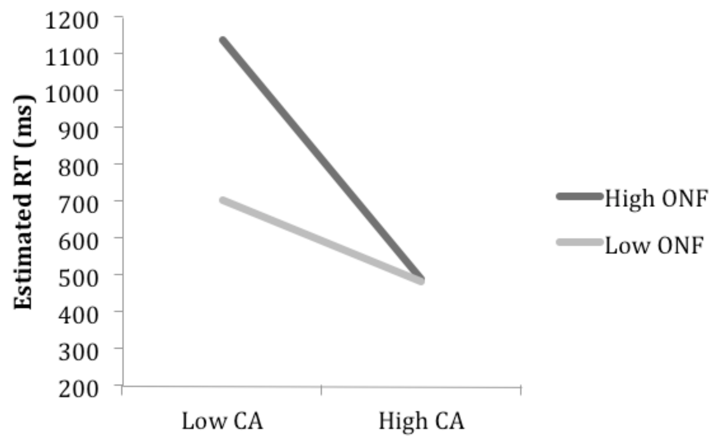

Figure 3. Estimated lexical decision latencies for high and low orthographic neighborhood frequency (ONF) and high and low context availability (CA)

4.0 DISCUSSION

A great deal of research has examined lexical and semantic variables that affect initial lexical access and how these variables interact. Early models of lexical access attempted to identify the individual variables that were most influential, but later models have acknowledged that many factors are at play, and instead focused on identifying interactions and changes over the time course of processing. However, the complexity of the literature in this area, the number of variables implicated, and the various study designs that have been used all make it difficult to be confident that any one study can tell the entire story, or even an accurate part of the story. In short, the current state of the literature is such that there is more disagreement than agreement concerning the key variables in lexical access.

The goal of the present study was to provide a more inclusive account of the variables that affect lexical decision, uniting for the first time a number of potential variables of interest that other researchers have identified. This has recently become possible due to advances in statistical methods, such as linear mixed effects modeling (Baayen et al., 2008). These models are capable of handling a large number of continuous variables, eliminating data loss due to dichotomization of variables as used in a majority of previous studies. In addition, we aimed to reexamine the findings of Tokowicz and Kroll (2007) in which the effects of number of meanings and NOS were conflated, and further improve on previous studies by using contextual diversity, a newer measure of word frequency that is more predictive of lexical decision performance than the Kucera and Francis (1967) norms (Adelman et al., 2006; Brysbaert & New, 2009; Perea et al., 2013; van Heuven, Mandera, Keuleers, & Brysbaert, 2014).

In the present study we examined a complex pattern of interactions between context availability, contextual diversity, polysemy, and orthographic neighborhood frequency. We first report several findings that are not surprising in light of past research, but documenting their replication in a new sample and using new statistical models is valuable given contradictory reports of how concreteness effects manifest during lexical decision, as reviewed in the Introduction. Then, we turn to discuss the novel patterns observed in these data.

4.1 RELATIONSHIP TO PAST FINDINGS

First, similar to reports by Rodd et al. (2002) and Beretta et al. (2005) we report a polysemy advantage in which words with a greater NOS are recognized more quickly than words with fewer senses. The stimuli used in this experiment were drawn from Tokowicz and Kroll (2007), and originally comprised both homonyms and polysemes. To verify that the ambiguity advantage reported in that study was due to polysemy, we removed all homonyms from the stimulus set and used only polysemes in our experiment.

Second, we found a significant main effect of contextual diversity, such that words that appear in a greater number of contexts are recognized more quickly than words that appear in fewer contexts. This is consistent with reports by Adelman et al. (2006) and Perea et al. (2013), and with the word frequency effect (Broadbent, 1967), although as we note in the results section contextual diversity is more predictive of lexical decision latencies than traditional word frequency measures (e.g., Adelman et al., 2006).

Third, we report a context availability advantage, in that words higher in context availability are recognized significantly more quickly than words lower in context availability. Further, with context availability controlled there is no effect of concreteness. These effects have previously been reported and add support to the context availability hypothesis (e.g., Schwaneflugel et al., 1988; Schwanenflugel & Shoben, 1983; Tokowicz & Kroll, 2007). This context availability advantage is at odds with the predictions of the DRV, because our study did not control for imageability, but nevertheless there was no effect of concreteness. We interpret these results as discouraging the application of dual-coding theory to lexical decision.

Fourth, we note two previously-documented effects of orthographic neighborhood in lexical decision. First, words with a greater number of higher-frequency neighbors are recognized more slowly than words with fewer higher-frequency neighbors. This neighborhood frequency effect has been reported by Grainger (1990) and Grainger, O'Reagan, Jacobs, and Segui (1992), and is thought to reflect the inhibition resulting from the activation of lexical patterns that are in competition with the target word. Second, we report an interaction between context availability and orthographic neighborhood frequency. Words higher in context availability are able to withstand the disadvantage of higher frequency orthographic neighbors, whereas words lower in context availability are slower to be recognized when they have a higher number of higher frequency orthographic neighbors. Put another way, words low in context availability are especially vulnerable to inhibition from higher frequency orthographic neighbors. This inhibition might delay the recognition of a word, and during this delay the reader attempts to access semantic information. Words that are low in semantic information (i.e., words that are low in context availability) are especially disadvantaged in this case. This finding is consistent with findings from Samson and Pillon (2014; Figure 1), although in that study concreteness was used instead of context availability. However, these variables are so highly correlated that finding the same interaction with context availability is not surprising.

We also report several findings that may be considered surprising in light of past research. First, we did not find evidence for the typically-observed concrete-word advantage (James, 1975; Kroll & Merves, 1986; Schwaneflugel et al., 1988; Schwanenflugel & Shoben, 1983), but rather found that higher concreteness was associated with a disadvantage in word recognition time. This is neither unprecedented nor unexpected. Although the concreteness advantage in lexical decision is often thought of as an ubiquitous effect, in reality it is not as canonical as thought – whereas some studies report a concreteness advantage (James, 1975), others report null or mixed findings (Kroll & Merves, 1986; Schwanenflugel & Shoben, 1983; Tolentino & Tokowicz, 2009), and still others report a reversal of the concreteness effect (Vigliocco, Kousta, Vinson, Andrews, & Del Campo, 2013).

This lack of consistency may seem puzzling, but our findings align with other studies reporting the concreteness effect was confounded by another variable. In support of this, we noted that with the effects of concreteness controlled, there is a reaction time advantage for words higher in context availability, and with the effects of context availability controlled there is a reaction time disadvantage for words higher in concreteness. This suggests that previous reports of the concrete-word advantage are actually reporting an advantage due to context availability, which has previously been proposed (Schwanenflugel et al., 1988; Schwanenflugel & Shoben, 1983; van Hell & de Groot, 1998b).

A second somewhat surprising finding is that we report an interaction between NOS and contextual diversity, in which the effects of NOS are most apparent for words low in contextual diversity. That is, words with fewer senses have an extra disadvantage when they are also used in fewer diverse contexts (e.g., words such as *stapler*, *tuba*, and *mittens*). Words with a greater amount of contextual diversity are already recognized quickly, and so are less affected by differing numbers of senses. In contrast, words that are used in fewer diverse contexts seem to be especially vulnerable to the effects of also being low in NOS. Although we are the first to report an interaction of polysemy and contextual diversity, we consider it only somewhat surprising because it is largely consistent with previous interactions between ambiguity and word frequency. For instance, Rubenstein et al. (1970) reported a significant interaction between ambiguity (nonhomographs vs. homographs) and word frequency (from the Lorge Magazine Count) in which the effects of ambiguity were significantly greater for low frequency words than high frequency words. Jaztrzembski (1981) and Pexman et al. (2004) both replicated this pattern with homonyms instead of homographs and used the Kucera and Francis (1967) frequency counts.

4.2 NOVEL RESULTS

We will now turn to discussing the novel results of our study. First, based on prior research, we were surprised that we did not find an interaction of NOS and context availability/concreteness. Tokowicz and Kroll (2007) reported that an ambiguity effect (specifically, a reaction time advantage for ambiguous words relative to unambiguous words) was present only for words that are low in context availability. Similarly, Jager and Cleland (2014) found that the existence of a polysemy advantage was moderated by the set of words used: They did not find a polysemy advantage for concrete nouns, but did for a mixed set of abstract nouns, verbs, and adjectives.

But, in the present study, the interactions between NOS and context availability and NOS and concreteness both failed to reach significance. There are several possible explanations for this difference. The first is that the word lists used by Tokowicz and Kroll contained homonyms. In the present study these words were removed, so that only polysemous words with varying NOS were included. With homonymous words removed, we did not replicate the NOS interaction with context availability. The second is study design: Jager and Cleland dichotomized NOS into few senses vs. many senses, whereas our study kept this variable continuous to minimize information loss, as recommended by Tokowicz and Warren (2008). We also note that Jager and Cleland (2014) did not directly test for an interaction between polysemy and concreteness, but rather tested for a polysemy advantage in two separate words sets (concrete nouns vs.) a mixed set of abstract nouns, verbs, adjectives; the observed difference between these word sets might have been influenced by some other property than concreteness.

The second novel finding of our study was the significant interaction of contextual diversity and context availability. In this interaction, words that are used in many diverse contexts are not very sensitive to context availability, whereas words that appear in fewer diverse contexts are more sensitive to context availability. Specifically, words lower in context availability and in contextual diversity (e.g., words such as *allegory*, *folly,* and *earnestness*) are especially slow to be recognized in a lexical decision task. This makes intuitive sense because these are words that one encounters infrequently and for which it is hard to call a context to mind. One would probably then expect that words high in both contextual diversity and contextual availability would be recognized most quickly, but interestingly this is not the case.

Rather, words low in contextual diversity but high in context availability (e.g., words such as *zebra*, *volcano*, and *mustache*) are recognized most quickly in this interaction, even more quickly than words high in both contextual diversity and context availability (e.g., words such as *head*, *room*, and *money*). We think this might be because the limited number of contexts in which a word like *zebra* occurs reduces competition between biasing contexts while the greater ease of accessing contextual information makes these words more distinctive and easier to be called to mind.

The findings of this study have implications for models of concreteness/context availability effects. In particular, our findings are relevant to the context availability hypothesis (e.g., Schwanenflugel & Shoben, 1983), which claims that concreteness effects arise as a result of the greater ease of retrieving contextual information for concrete words over abstract words. That is, in situations in which it is difficult to decide if a letter string is a legal word, such as a lexical decision task in which some items are pseudohomophones, readers attempt to use contextual information to disambiguate the letter string. Readers are able to think of contexts in which a word can occur more quickly when words have more semantic information available, allowing them to resolve lexical ambiguity more quickly than for words that have lower context availability. The context availability hypothesis predicts that if the effects of context availability are controlled, the concreteness effect will disappear. We included both concreteness and context availability in our model and found that, as predicted by the context availability hypothesis, the concreteness advantage disappeared and a significant concreteness disadvantage emerged. Furthermore, with concreteness controlled there was a significant context availability advantage. These findings support the predictions of the context availability hypothesis.

Our results are also consistent with the predictions of the SSD account of semantic ambiguity resolution (Armstrong & Plaut, 2016), which predicts that words with a greater number of senses will have a processing advantage over words with a lower number of senses or unambiguous words. This advantage is thought to arise because the greater the number of related senses a word has the greater amount of excitatory feedback available during early processing, which drives a word towards a recognition threshold more quickly.

4.3 CONCLUSION

Overall, this research advances the study of semantic ambiguity and lexical access in three ways. First, it replicates the polysemy advantage and demonstrates that this advantage depends on contextual diversity. Second, this study highlights the importance of both orthographic neighborhood frequency and contextual diversity as lexical variables that interact with other lexical and semantic variables in the process of word recognition, and additionally supports the use of contextual diversity as a sensitive and predictive measure of lexical decision latencies. Finally, this study supports the context availability hypothesis and the SSD model of semantic ambiguity resolution.

The results of this study provide insight into the nature of the polysemy advantage and also provide direction for future research. Our study highlights the importance of considering the continuous nature of polysemy rather than classifying polysemous words into categories based on number of senses. However, recent research has suggested that there may be differential effects of ambiguity even within polysemous words. For instance, metonyms are known to be processed more quickly than metaphors (Eddington & Tokowicz, 2015), but thus far no one has examined whether these different types of polysemy interact differently with context availability or contextual diversity. Future research should take this into account and further explore how the results outlined in this paper may or may not be present across other tasks and throughout the time course of lexical processing.

5.0 FOOTNOTES

[1]Reported correlations of concreteness and imageability: Paivio (1966): $r = .77$; Paivio (1968): $r = .78$; Reilly and Kean (2007): $r = .90$; Rice et al. (present study): $r = .87$

[2] Reported correlations of concreteness and context availability: de Groot et al. (1994): $r = .82$; Schwanenflugel et al. (1988): $r = .69$; Jones (1985): $r = .88$. We noted that in our sample concreteness and context availability are highly correlated ($r = .76$). To verify that the two variables are not redundant, we conducted tests of model fit in which we compared Model 1 (which included concreteness and context availability; see Figure 2), to Model 1a (identical to Model 1 except context availability was dropped), and Model 1b (identical to Model 1 except concreteness was dropped) using likelihood ratio Chi-Square tests for nested models. We found that the model that included both concreteness and context availability (Model 1) was the best fitting model, which indicates that concreteness and context availability are capturing distinct portions of the variance in lexical decision latencies. Although this comparison shows that both variables together are better than either variable alone, we wanted to test whether the model with only context availability or the model with only concreteness model was a better fit. Therefore, we also compared Model 1a to Model 1b using Bayesian Information criterion (BIC) for non-nested models, and found that Model 1b (context availability only) was a significantly better fit for the data. Finally, we examined whether the choice to include concreteness or context availability, or both affected the overall pattern of results, the magnitude or direction of

parameter estimates, and most crucially if it impacted the critical hypothesized interaction (concreteness/CA and NOS). We found that the overall pattern of parameter estimates was the same, and the interaction of concreteness/CA was never significant. Therefore we conclude that collinearity between concreteness and context availability does not unduly influence our results.

[3] As described by Barr, Levy, Scheepers, and Tily (2013), a maximal random effects structure is preferred for hypothesis testing with linear mixed effects models. Given the complexity of our dataset, a maximal model failed to converge, and so we used a theory-guided approach to design our random effects structure. We first entered random slopes for the two variables we hypothesized would interact with NOS: context availability and contextual diversity. We then took a data-driven approach and used likelihood ratio tests to determine whether these inclusions resulted in a better-fitting model of the data, which we found to be true.

[4] We tested and excluded several alternative models. First, we examined whether word frequency might be a better predictor than contextual diversity, by substituting a zipf-transformed measure of word frequency from the eLexicon SUBTL-US corpus into the model instead of contextual diversity. Consistent with other comparisons of contextual diversity and word frequency (e.g., Adelman, 2006), model comparison revealed that the contextual diversity model fit better than the word frequency model. Bayesian Information Criterion (BIC) was used to compare the fit of the non-nested contextual diversity and the word frequency models. BIC is an estimate of the information lost from reducing a presumed 'true' underlying model to the existing model. A lower BIC value indicates that less information is lost, and thus shows which model is a better fit for the dataset. BIC for the contextual diversity model was 436,654, whereas BIC for the word frequency model was only 436,491, meaning that otherwise identical models were improved by using contextual diversity instead of word frequency. Second, a 3-way interaction of contextual

diversity, NOS, and concreteness was tested but failed to reach significance. Finally, we tested to see if concreteness/context availability would explain RT variance in the absence of contextual diversity. Running Model 1 without contextual diversity revealed significant main effects of concreteness (b = 16.19, SE = 1.97, t = 8.23) and context availability (b = -73.28, SE = 6.03, t = -12.16), as well as a significant interaction of context availability and NOS (b = 4.30, SE = 1.59, t = 2.70). When contextual diversity is present, the main effects of concreteness and context availability are still statistically significant but attenuated. However, the interaction of context availability are still significant but attenuated. However, the interaction of context availability and NOS is only significant when contextual diversity is not present. Therefore, we decided to use contextual diversity instead of traditional word frequency as it improves model fit.

APPENDIX A

Table 5. Concreteness and context availability ratings for stimuli

Stimulus	*Concreteness*	*Context availability*
ability	2.35	5.13
abuse	3.01	5.43
action	3.06	5.44
address	4.46	6.04
adult	5.42	6.36
advantage	2.20	5.24
advice	2.75	5.36
age	3.37	5.94
agility	2.94	5.13
airplane	6.51	6.39
allegory	2.18	3.34
altar	5.93	5.67
ambulance	6.48	6.31
anger	2.32	5.71
anguish	2.20	4.63
animal	6.11	6.50
answer	3.01	5.70
anxiety	2.49	5.27
apple	6.70	6.57
aptitude	2.00	4.00
army	5.65	6.31
arrival	3.21	5.41
arrow	6.10	6.29
art	4.01	6.23
artist	5.38	5.96
atrocity	2.17	4.09
attacks	3.82	5.74
attempt	2.77	5.01
attention	2.49	5.31
attitude	2.30	5.37
audience	5.21	6.04
avocado	6.70	6.26
awareness	2.27	4.87

baby	6.41	6.54
bald	5.63	6.10
bath	6.18	6.29
beauty	2.56	5.71
bed	6.49	6.53
beer	6.44	6.60
belief	2.04	5.40
bird	6.59	6.57
blame	2.00	4.99
blanket	6.42	6.39
blessing	2.13	5.10
blind	3.99	6.00
body	5.79	6.26
boil	4.80	5.97
bone	6.51	6.37
book	6.34	6.44
boot	6.68	6.31
boredom	2.21	5.86
bottle	6.30	6.33
box	5.99	6.21
boys	6.07	6.59
bread	6.63	6.41
break	3.63	5.49
brick	6.39	6.37
buffalo	6.39	6.29
bullet	6.34	6.33
bungalow	5.62	4.53
butter	6.51	6.31
button	6.28	6.29
buy	3.37	5.84
calendar	6.08	6.29
candle	6.42	6.30
capacity	2.73	4.49
capitol	4.65	5.67
captain	5.38	6.20
car	6.61	6.20
carrot	6.82	6.26
carry	3.79	5.57
cart	6.17	6.09
case	4.89	5.51
cat	6.55	6.54
cause	2.49	4.93
century	3.52	5.07
chair	6.44	6.41
chalk	6.44	6.27
chaos	2.42	5.37

childhood	3.17	6.00
children	5.96	6.39
church	5.76	6.29
cigar	6.51	6.30
circle	5.45	6.26
clothes	6.31	6.39
coast	5.27	6.01
coat	6.37	6.26
coffee	6.73	6.51
coins	6.41	6.31
color	4.54	6.23
comb	6.44	6.23
comedy	2.94	5.90
compulsion	2.25	4.40
computer	6.48	6.59
concept	1.69	4.39
condition	2.63	5.10
confession	2.61	5.46
cooks	5.42	6.13
corn	6.38	6.26
corner	5.20	6.09
cotton	6.34	6.11
cow	6.55	6.36
create	2.49	5.21
creation	2.49	4.96
crisis	2.45	5.43
cruelty	2.41	5.23
culture	2.86	5.61
cup	6.59	6.51
custom	2.66	5.03
danger	2.73	5.73
daring	2.17	5.06
data	4.51	5.74
daughter	5.59	6.21
death	3.35	6.00
deceit	2.20	4.57
decency	1.92	4.49
decision	2.44	5.24
demand	2.42	5.30
demon	3.31	5.80
desk	6.42	6.43
detail	2.97	5.09
devil	2.90	5.99
devotion	2.41	4.96
difference	2.49	5.27
difficulty	2.41	5.39

dinner	5.75	6.39
director	5.39	6.03
discover	2.17	5.04
disorder	2.77	5.27
district	3.68	4.93
divided	2.92	5.41
doctor	6.27	6.63
dollar	5.75	6.33
door	6.52	6.33
dream	2.61	5.66
dress	6.30	6.50
dry	4.30	5.91
earnestness	1.96	3.54
ease	2.08	4.84
economy	3.24	5.47
effort	2.21	5.03
elbow	6.31	6.33
elephant	6.72	6.57
elevator	6.39	6.37
elm	5.85	4.31
enemy	3.63	5.79
English	4.00	6.07
equity	2.25	3.86
example	3.00	5.03
expression	2.63	5.29
eye	6.45	6.40
face	5.94	6.13
fact	3.46	5.64
factory	6.14	6.10
faith	1.97	5.39
fame	2.42	5.70
family	5.13	6.39
fantasy	2.18	5.73
farmer	6.14	6.24
fate	1.77	4.74
favor	2.52	5.21
feather	6.31	6.23
feud	2.80	4.96
fever	4.76	6.04
finger	6.80	6.37
fire	6.03	6.36
fist	6.08	6.13
flame	5.87	6.17
flaw	2.87	5.29
floor	6.42	6.20
flour	6.55	6.11

flower	6.46	6.57
fly	5.58	6.21
folly	2.46	3.43
foot	6.38	6.46
forest	6.32	6.43
freedom	2.13	5.80
fruit	6.45	6.64
fun	2.72	5.74
furniture	6.51	6.29
future	2.17	5.61
game	4.76	6.37
garden	6.30	6.46
gathering	3.79	5.33
ghost	3.37	6.24
gift	5.30	6.19
glory	1.80	5.06
glue	6.23	6.06
gold	5.85	6.24
governor	5.65	6.03
grandmother	6.38	6.69
grapes	6.63	6.46
grass	6.44	6.40
gray	4.63	6.04
greed	2.25	5.50
green	4.83	6.27
grief	2.25	5.41
guess	2.48	5.11
guitar	6.56	6.47
hair	6.52	6.64
happy	2.35	5.97
hat	6.48	6.33
hatred	2.01	5.53
head	6.45	6.33
health	2.97	5.57
heart	5.44	6.43
heaven	2.14	5.64
hell	2.31	5.69
highway	6.27	6.40
honey	6.06	6.24
honor	2.35	5.11
hope	2.11	5.31
horse	6.61	6.27
hospital	6.35	6.57
hotel	6.32	6.34
hour	3.54	5.96
house	6.42	6.67

hunger	3.65	6.11
idea	2.20	5.29
illusion	2.68	4.96
impression	2.11	5.36
Indians	5.90	5.97
influence	2.28	5.29
inheritance	3.11	5.39
innocence	2.03	5.20
insect	6.23	6.16
inside	3.83	5.46
insight	2.32	4.79
insult	2.82	5.37
irony	1.79	4.63
jeopardy	2.85	5.37
joke	3.04	6.09
joy	2.38	5.67
juice	6.32	6.43
jungle	6.00	6.13
justice	2.34	5.34
keys	6.56	6.37
kindness	2.49	5.54
kitchen	6.20	6.41
lab	5.62	6.24
labyrinth	4.61	4.84
lady	5.86	6.14
lake	6.32	6.36
laugh	4.31	6.09
lawyer	6.06	6.11
leaf	6.58	6.46
leave	3.03	5.21
leg	6.51	6.44
lemon	6.55	6.43
letter	5.87	6.10
library	6.06	6.37
lie	2.34	5.63
light	4.66	5.99
loan	3.68	5.37
lock	5.82	6.20
love	2.39	5.97
loyalty	1.90	5.31
luck	2.10	5.56
luxury	2.89	5.54
machine	5.77	6.13
madness	2.01	5.41
mail	5.94	6.11
malice	2.20	3.60

manner	2.17	4.73
map	6.00	6.13
mark	4.30	5.41
market	5.54	6.19
married	3.39	6.24
material	4.83	5.46
mayor	5.96	6.27
meaning	1.61	4.31
meat	6.21	6.31
mechanic	5.55	5.91
memory	2.51	5.49
men	5.79	6.23
mercy	1.99	5.03
method	2.63	4.93
middle	3.69	5.79
mind	2.89	5.50
minute	3.41	5.81
miracle	2.28	5.49
mirror	6.20	6.26
mischief	2.35	5.27
misery	2.27	5.31
missile	6.17	5.87
mistakes	2.52	5.30
mittens	6.37	6.17
moment	2.35	4.96
money	5.58	6.36
moral	1.97	5.10
mother	5.70	6.39
mountain	6.46	6.50
mouse	6.76	6.36
moustache	6.44	6.20
movie	5.86	6.54
mud	6.35	6.16
music	4.90	6.29
mustache	6.59	6.51
necklace	6.59	6.33
needle	6.42	6.27
newspaper	6.51	6.43
night	4.62	6.16
noise	4.30	5.81
nonsense	2.23	4.91
norm	2.15	4.69
nucleus	4.72	5.73
nurse	6.28	6.39
oath	2.70	5.16
office	5.73	6.31

officer	5.92	6.24
old	3.15	6.03
onion	6.63	6.29
opera	5.59	6.09
operator	5.51	5.61
opinion	2.21	5.36
organdy	3.80	1.43
owl	6.75	6.19
pain	3.45	5.87
pamphlet	6.42	5.47
panic	2.38	5.64
panther	6.73	6.66
paper	6.42	6.44
park	5.85	6.36
passion	2.56	5.57
patience	2.14	5.37
patients	5.51	6.19
peace	2.11	5.64
pearl	6.38	6.07
pencil	6.45	6.43
people	5.65	6.07
pepper	6.45	6.13
person	5.80	6.01
photo	6.01	6.23
phrases	3.35	5.09
pianist	6.21	6.03
picnic	5.51	6.17
pinnacle	3.34	4.11
pirate	5.99	6.30
plane	6.30	6.37
plate	6.24	6.26
points	3.38	5.64
policeman	6.38	6.59
pope	5.97	6.01
pound	4.68	5.74
power	2.37	5.54
preference	1.79	4.54
prestige	2.15	4.27
price	3.83	6.03
pride	2.32	5.47
princess	5.75	6.49
prize	4.92	6.07
promise	2.44	5.50
pumpkins	6.59	6.61
rabbit	6.56	6.41
rage	2.54	5.36

rain	6.15	6.31
raisin	6.70	6.13
reaction	2.83	5.19
reason	2.20	5.03
recipe	4.96	6.03
red	4.72	6.34
remember	2.38	5.24
remnant	3.20	4.24
repair	3.35	5.54
resistance	2.75	4.74
respect	2.14	5.49
rhythm	3.34	5.41
rice	6.62	6.50
river	6.20	6.34
room	5.55	6.23
rose	6.49	6.60
rug	6.35	6.20
run	4.49	6.14
sadness	2.34	5.37
safety	2.48	5.67
sailor	6.11	6.24
satire	2.48	4.76
school	5.80	6.51
scissors	6.59	6.54
sea	6.10	6.37
seconds	3.58	5.96
sections	4.01	5.07
shame	2.01	5.01
sheep	6.63	6.31
shirt	6.55	6.56
shoe	6.58	6.49
shop	5.08	6.14
shoulder	6.30	6.23
sidewalk	6.42	6.21
silence	3.39	5.57
silver	5.56	6.07
sin	2.32	5.81
situation	2.83	4.89
size	3.42	5.47
skirt	6.45	6.21
smoke	5.69	6.26
snow	6.17	6.50
soap	6.41	6.21
song	4.77	6.21
soul	2.14	5.43
soup	6.54	6.31

south	3.18	5.77
space	3.23	5.70
spell	2.97	5.69
spider	6.49	6.49
spirit	2.00	5.14
spoon	6.42	6.39
square	5.28	6.27
stalk	3.89	5.26
stapler	6.39	6.36
station	5.24	5.89
stench	4.27	5.47
stone	6.18	6.14
story	4.20	5.66
strawberry	6.82	6.46
street	5.94	6.40
strength	3.03	5.60
string	6.17	6.23
struggle	2.73	5.23
student	5.62	6.26
stuff	3.87	4.89
sugar	6.37	6.43
summer	4.18	6.37
sun	6.25	6.51
supper	5.46	6.20
support	2.79	5.31
sweat	5.99	6.36
table	6.54	6.44
talent	2.69	5.53
task	3.69	5.46
tea	6.39	6.50
telephone	6.37	6.46
temerity	2.13	1.93
tendency	2.20	4.51
terrain	5.17	5.09
theater	5.93	6.24
theme	2.75	5.17
theory	1.75	4.93
thing	3.59	4.73
thought	2.21	5.39
thread	5.97	5.94
threat	2.52	5.51
throat	6.41	6.14
thumb	6.59	6.13
tiger	6.77	6.39
tobacco	6.37	6.20
tomato	6.70	6.46

tooth	6.58	6.44
tower	6.07	6.19
town	5.35	6.11
tragedy	2.37	5.64
train	6.32	6.33
tranquility	2.20	4.71
trauma	2.85	5.51
tray	6.44	5.97
tree	6.63	6.40
triangle	5.10	5.91
truck	6.49	6.29
truth	2.08	5.51
tuba	6.52	6.20
turkey	6.70	6.43
twins	6.04	6.40
umbrella	6.65	6.59
vanity	2.65	4.37
vegetable	6.27	6.40
victim	4.32	5.89
victory	2.65	5.97
view	3.25	5.46
village	5.66	6.17
violence	3.24	5.80
violin	6.58	6.43
virtue	1.86	4.20
virus	4.58	5.94
voice	4.72	6.03
volcano	6.54	6.37
wall	6.06	6.34
war	3.99	6.26
warmth	4.39	6.09
watch	5.75	6.29
water	6.56	6.50
wave	5.52	6.34
weakness	2.32	5.31
week	3.44	5.97
welfare	2.76	5.14
whale	6.54	6.37
wheat	6.34	6.19
wheel	6.39	6.31
window	6.14	6.34
winter	4.83	6.19
woman	5.89	6.37
word	4.06	5.90
work	4.15	5.99
wrinkle	5.25	6.01

year	3.62	5.80
yesterday	3.52	5.67
youth	3.30	5.54
zebra	6.72	6.34

BIBLIOGRAPHY

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814-823.

Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, CA: Sage Publications.

Andrews, S. (1989). Frequency and neighborhood size effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 802-814.

Armstrong, B. C., & Plaut, D. C. (2008). Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.) *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 273-278). Austin, TX: Cognitive Science Society.

Armstrong, B. C., & Plaut, D. C. (2011). *Inducing homonymy effects via stimulus quality and (not) nonword difficulty: Implications for models of semantic ambiguity and word recognition.* Paper presented at the Proceedings of the 33rd Annual Conference of the Cognitive Science Society, Austin, TX.

Armstrong, B. C., & Plaut, D. C. (2016). *Semantic ambiguity effects in lexical procecssing: A neural-network account based on semantic settling dynamics*. Unpublished manuscript, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA.

Azuma, T., & van Orden, G. C. (1997). Why SAFE Is Better Than FAST: The relatedness of a word's meanings affects lexical decision times. *Journal of Memory and Language, 36*, 484-504.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York, NY: Cambridge University Press.

Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon, 5*(3), 436-461. doi: 10.1075/ml.5.3.10baa

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. doi: 10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3). doi: 10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (R Package version 1.1-7). Retrieved from http://CRAN.R-project.org/package=lme4

Begg, I., & Paivio, A. (1969). Concreteness and imagery in sentence meaning. *Journal of Verbal Learning and Verbal Behavior, 8*, 821-827.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: an MEG study. *Cognitive Brain Research, 24*, 57-65. doi: 10.1016/j.cogbrainres.2004.12.006

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review, 74*(1), 1-15.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavioral Research Methods, 41*(4), 977-990. doi: 10.3758/BRM.41.4.977

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256.

de Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 824-845.

Eddington, C. M., & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: the current state of the literature. *Psychonomic Bulletein Review, 22*(1), 13-37. doi: 10.3758/s13423-014-0665-7

Ferraro, F. R., & Hansen, C. L. (2002). Orthographic neighborhood size, number of word meanings, and number of higher frequency neighbors. *Brain & Language, 82*(2), 200-205.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*, 627-635.

Frank, A. (2011). Diagnosing collinearity in mixed models from lme4. HLP/Jaeger lab blog. Retrieved from https://hlplab.wordpress.com/2011/02/24/diagnosing-collinearity-in-lme4/.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language, 29*, 228-244.

Grainger, J., O'Reagan, J. K., Jacobs, A. M., & Segui, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception and Psychophysics, 51*(1), 49-56.

Hino, Y., Pexman, P. M., & Lupker, S. J. (2006). Ambiguity and relatedness effects in semantic tasks: Are they due to semantic coding? *Journal of Memory and Language*, *55*(2), 247-273. doi: 10.1016/j.jml.2006.04.001

Jager, B., & Cleland, A. A. (2014). Polysemy Advantage with Abstract But Not Concrete Words. *Journal of Psycholinguistic Research, 45*, 143-156. doi: 10.1007/s10936-014-9337-z

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 104*(2), 130-136.

Jaztrzembski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology, 13*, 278-305.

Jones, G. V. (1985). Deep dyslexia, imageability, and ease of prediction. *Brain & Language, 24*, 1-19.

Klein, D. E., & Murphy, G. L. (2001). Paper has been my ruin: conceptual relations of polysemous senses. *Journal of Memory and Language, 47*, 548-570.

Klepousniotou, E. (2002). The processing of lexical ambiguity: homonymy and polysemy in the mental lexicon. *Brain & Language, 81*(1-3), 205-223.

Kroll, J. F., & Merves, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning Memory and Cognition, 12*, 92-107.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: 1. An account of basic findings. *Psychological Review, 88*(1), 375-407.

Millis, M. L., & Button, S. B. (1989). The effect of polysemy on lexical decision time: now you see it, now you don't. *Memory & Cognition, 17*(2), 141-147.

Mirman, D., Strauss, T. J., Dixon, J. A., & Magnuson, J. S. (2010). Effect of representational distance between meanings on recognition of ambiguous spoken words. *Cognitive Science, 34*(1), 161-173.

Paivio, A. (1968). A factor-analytic study of word attributes and verbal learning. *Journal of Verbal Learning and Verbal Behavior, 7*, 41-49.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinchart, and Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.

Perea, M., Soares, A. P., & Comesana, M. (2013). Contextual diversity is a main determinant of word identification times in young readers. *Journal of Experimental Child Psychology, 116*(1), 37-44. doi: 10.1016/j.jecp.2012.10.014

Pexman, P. M., Hino, Y., & Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1252-1270. doi: 10.1037/0278-7393.30.6.1252

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*(5), 377-500. doi: 10.1080/02643299308253469

R Core Team (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125-157.

Reilly, J., & Kean, J. (2007). Formal distinctiveness of high- and low-imageability nouns: analyses and theoretical implications. *Cognitive Science, 31*(1), 157-168. doi: 10.1080/03640210709336988

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making Sense of Semantic Ambiguity: Semantic Competition in Lexical Access. *Journal of Memory and Language, 46*(2), 245-266. doi: 10.1006/jmla.2001.2810

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9*, 487-494.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89*(1), 60-94.

Samson, D., & Pillon, A. (2004). Orthographic neighborhood and concreteness effects in the lexical decision task. *Brain & Language, 91*(2), 252-264. doi: 10.1016/j.bandl.2004.03.001

Schwaneflugel, P. J., Harnishfeger, P. J., & Stowe, R. W. (1988). Context Availability and Lexical Decisions for Abstract and Concrete Words. *Journal of Memory and Language, 27,* 499-520.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 82-102.

Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly, 24,* 114-126. doi:10.3758/BF03208259

Spreen, O., & Shulz, R. W. (1966). Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. *Journal of Verbal Learning and Verbal Behavior*, *5*, 459-468.

Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes, 22*(5), 727-779. doi: 10.1080/01690960601057068

Tokowicz, N., & Warren, T. (2008). Quantification and statistics. In L. Wei & M. Moyer (Eds.), *The Blackwell guide to research methods in bilingualism and multilingualism* (pp. 214-231). Malden, MA: Blackwell Publishers.

Tolentino, L. C., & Tokowicz, N. (2009). Are pumpkins better than heaven? An ERP investigation of order effects in the concrete-word advantage. *Brain & Language, 110*(1), 12-22. doi: 10.1016/j.bandl.2009.01.001

van Hell, J. G., & de Groot, A. M. (1998a). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*(2), 193-211. doi: 10.1017/

van Hell, J. G., & de Groot, A. M. (1998b). Disentangling context availability and concreteness in lexical decision and word translation. *Quarterly Journal of Experimental Psychology, 51A,* 41-63.

van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: a new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176-1190. doi: 10.1080/17470218.2013.850521

Vigliocco, G., Kousta, S., Vinson, D., Andrews, M., & Del Campo, E. (2013). The representation of abstract words: what matters? Reply to Paivio's (2013) comment on Kousta et al. (2011). *Journal of Experimental Psychology General, 142*(1), 288-291. doi: 10.1037/a0028749

Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: the role of emotion. *Cereb Cortex, 24*(7), 1767-1777. doi: 10.1093/cercor/bht025