THE EFFECTS OF MEANING DOMINANCE AND MEANING RELATEDNESS ON AMBIGUITY RESOLUTION: IDIOMS AND AMBIGUOUS WORDS

by

Evelyn Arko Milburn

B.A., University of California, Davis, 2010

M.S., University of Pittsburgh, 2014

Submitted to the Graduate Faculty of the

Kenneth P. Deitrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

KENNTH P. DEITRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Evelyn Arko Milburn

It was defended on

November 21, 2017

and approved by

Michael Walsh Dickey, Associate Professor, Departments of Communication Sciences and Disorders and Psychology

Charles Perfetti, Distinguished University Professor, Department of Psychology

Natasha Tokowicz, Associate Professor, Departments of Psychology and Linguistics

Dissertation Advisor: Tessa Warren, Associate Professor, Departments of Psychology, Linguistics, and Communication Sciences and Disorders

THE EFFECTS OF MEANING DOMINANCE AND MEANING RELATEDNESS ON AMBIGUITY RESOLUTION: IDIOMS AND AMBIGUOUS WORDS

Evelyn Arko Milburn, PhD

University of Pittsburgh, 2017

Figurative language is language in which combining the meanings of the individual words in an expression leads to a different meaning than the speaker intends (Glucksberg, 1991), resulting in potential ambiguity between meanings. In this dissertation, we tested the predictions of a sentence processing framework in which literal and figurative language are not truly distinct. To do this, we examined the effects of two constructs—meaning dominance and meaning relatedness—on comprehension of idioms and ambiguous words. Processing similarities between these two types of ambiguous unit would indicate that ambiguities are resolved using the same processes during language comprehension, and therefore that literal and figurative language are broadly similar rather than being categorically distinct. In two parallel sub-experiments, Experiment 1 compared facilitation for dominant and subordinate meanings of ambiguous units in a primed lexical decision task. For ambiguous words, participants showed greater facilitation when one meaning was strongly dominant. For idioms, participants showed greater facilitation for idioms compared to control phrases, and lowest accuracy when responding to literal target words following highly figuratively-dominant idioms. Experiment 2 used eyetracking during reading to examine how biasing context affected idiom meaning activation, as well as how idiom meanings were integrated into a larger text. Participants read the idioms slowest when both figurative dominance and meaning relatedness were high, and fastest when meaning relatedness was high and figurative dominance was low, replicating results for ambiguous word reading found by Foraker and Murphy (2012). This is suggestive evidence for a language comprehension system that resolves ambiguities similarly regardless of grain size or literality. We also found facilitative effects of meaning relatedness in idiom reading parallel to the polysemy advantage in ambiguous word research, providing evidence that meaning relatedness is universal to many types of ambiguity resolution. The present study provides preliminary evidence that idioms and ambiguous words are treated similarly during ambiguity resolution. These results have implications for our understanding of idiom comprehension, and suggest valuable new avenues for future research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Thanks first and foremost to my academic advisor, Tessa Warren. Tessa, thank you for teaching me to write and to think, for your unwavering support, and for always challenging me to be a better researcher. Thanks also for fantasy novel recommendations, vegetarian cookbooks, and the most delicious pumpkin pie.

Thanks to Michael Walsh Dickey, who might as well have been my unofficial secondary advisor. Mike, thank you for your help and support for all these years. Can't stop the science!

Thanks to my other dissertation committee members, Natasha Tokowicz and Charles Perfetti, for their helpful suggestions. Thanks also to Scott Fraundorf for statistical advice and endless patience.

Thanks to my parents, Karen Arko and Al Milburn. Mom, thank you for believing in me, for letting me do my own thing, and for always being ready to talk about knitting. Dad, thank you for making me laugh and singing me songs, for reminding me to not take myself too seriously, and for rainy walks around Phoenix Lake. Thanks also to my sister, Sarah Liebowitz. Sarah, thank you for always understanding me, for being my tattoo buddy, for being the best big sister a girl could ever want.

Thanks to my friends and fellow students. Michelle, thank you for being a truly awesome labmate and friend (and thanks for providing mini Heath bars when I needed them. I'm eating

one right now). Katherine, thank you for your unparalleled friendship, your intelligence, and for fighting with R together in Coffee Tree. Alba, thank you for “Alba hugs”, getting me into rowing, and supporting me during these last few weeks of dissertation writing. Chelsea, thank you for your friendship, your style, and your 2015 review paper that provided the germ of inspiration for this project. I literally couldn’t have done this without you. Joe, thank you for your humor, your friendship, and for taking excellent care of my cat while I’m away (thanks also to Joce, for the same things and for occasionally cleaning my apartment unprovoked). Becca, thank you for being my accountability buddy, for being a giant nerd with me, and for standing for hours to see the Mountain Goats. Mehrgol, thank you for being an excellent RA turned excellent BPhil student turned excellent friend, and for your all-around excellence. My-Hai and Steph, thank you for being my rocks, my voices of non-academic (mostly) sanity, for always cheering me on.

Thanks to my fellow Cognitive graduate students, for mutual support and commiseration (Latin, from *com* + *miserari*, “bewail, lament”).

Thanks to my cat, Miriel Miel, for sleeping on my feet every night, for a purr that sounds like a Geiger counter, and for taking perfect simple joy in a piece of string.

Finally, thanks to the many research assistants who have helped me with projects over the years, the LABlings of the LABlab, and my participants.

1.0 INTRODUCTION

Figurative language is language in which combining the meanings of the individual words in an expression leads to a different meaning than the speaker intends (Glucksberg, 1991). Types of figurative language include idiom, metaphor, hyperbole, and irony, among others: all of these types of expression involve a discrepancy between the literal words that are said and the figurative meaning that is intended. Much previous research on figurative language comprehension has focused on accounting for the differences between literal and figurative language. This research has been critical for building a picture of how figurative language may be processed, but this perspective has caused less attention to be paid to potential similarities between literal and figurative language processing. However, more recent research has shown parallels between processing of literal and figurative language (Cutting & Bock, 1997; Giora, 2002; Konopka & Bock, 2009), suggesting that considering similarities between these two apparently distinct forms of language can yield critical new knowledge about language processing as a whole.

In light of this, the overarching goal of the present research is to investigate whether or not the same mechanisms underlie literal and figurative ambiguity resolution. We predict that robust patterns of processing in literal language—specifically, patterns related to the processing of ambiguous words—will also be found in processing of ambiguous figurative units such as idioms. Such findings would indicate that ambiguities are resolved using the same processes during language comprehension, and therefore that literal and figurative language, rather than being categorically distinct, are instead broadly similar.

There is growing evidence that common strings of words—frequently referred to as "multiword phrases"—have effects on processing similar to effects of single words, suggesting a flexible language system that is able to represent and process single words and multiword units simultaneously. Comprehenders are sensitive to the frequencies of multiword phrases such that more frequent literal phrases are processed more quickly (Arnon & Snider, 2010; Siyanova-Chanturia, Conklin, & van Heuven, 2011) and remembered more accurately (Tremblay, Derwing, Libben, & Westbury, 2011) than less frequent phrases. These results suggest thatat least some multiword phrases are psychologically salient, and may be processed as whole units in a "word-like" manner.

This tension between word-level and phrase-level meanings is present in figurative as well as literal language, and it has driven the creation of several models of idiom representation. Older models typically posit that idioms are represented in the lexicon as single words (Swinney & Cutler, 1979), whereas newer models are more likely to represent idiom processing as at least partially compositional (Titone & Connine, 1999). However, this rigid dichotomy between lexicalized and compositional idiom representation may also be artificial. If the language system is sensitive to both single word and multiword units—as supported by evidence from multiword phrase processing—idioms may also be represented in a more wordlike manner while still showing effects of their component words on processing. Under this view, multiword phrases, whether literal or figurative, are treated the same by the language processing system. This means that characteristics influencing single word processing should also influence processing of multiword phrases.

The literature on ambiguous word processing is a potentially fruitful source of characteristics that may influence idiom processing. Conceptualizing idioms as ambiguous units is not a new idea (Cronk, Lima, & Schweigert, 1993), and studies of idiom comprehension frequently implicitly draw on concepts from the ambiguous word processing literature to make predictions about idiom processing. One goal of the present research is to explicitly compare processing of idioms and ambiguous words in a way that thus far has not been attempted in idiom research.

One characteristic of ambiguous words that may also influence idiom processing is the degree of relatedness between a word's meanings. Ambiguous words can be categorized as homonyms or polysemes depending on whether or not their meanings are semantically related. Similarly, the literal and figurative meanings of idioms can also be more or less semantically related to each other. Some research has shown a processing advantage for idioms with more related meanings (Caillies & Butcher, 2007; Titone & Connine, 1999), but different operationalizations of meaning relatedness across studies make results difficult to compare. Another established characteristic of ambiguous words is meaning dominance, in which one meaning of a word is more commonly used or easily accessed than another. This characteristic is also shared by idioms: figurative meanings of some idioms are dominant and easily accessed even in isolation (Gibbs, 1980), whereas other figurative expressions may have meanings that are more balanced between the literal and figurative. Meaning dominance and meaning relatedness interact during processing of ambiguous words (e.g., Foraker & Murphy, 2012). Finding similar interactions in idioms would be suggestive evidence that literal and figurative language are processed using the same mechanisms, and that the properties of meaning dominance and meaning relatedness influence processing of units of language larger than single words.

This research is intended to test the predictions of a sentence processing framework in which literal and figurative language are not truly distinct. Under this view, language input at multiple grain sizes is processed simultaneously, and multiple meaning mappings are typical for both words and phrases; the same processing mechanisms are used for wordlike units, regardless of whether they are single words or multiword chunks, leading to similar processing effects for ambiguous words and idioms.

If literal and figurative language are not categorically distinct, and if single words and phrases are processed in the same ways, then similar constructs should have the same effects on processing of both idioms and ambiguous words. We predict that meaning relatedness (in ambiguous words) and transparency (in idioms) are essentially the same construct, and therefore will have similar effects on processing of idioms and ambiguous words. Second, we predict that idioms and ambiguous words will show parallel effects of meaning dominance. In particular, we predict that meaning dominance and meaning relatedness will interact to drive processing similarly in idioms and ambiguous words. A final goal of this work is to evaluate models of both idiom processing and ambiguous word processing based on their ability to accommodate the results of the experiments in the present study.

1.1 BACKGROUND

Situating Idioms Within Language Comprehension

There are roughly three different types of models of idiom representation (Libben & Titone, 2008): *noncompositional* models such as Swinney and Cutler's Lexical Representation Hypothesis (1979) or Gibbs's Direct Access Model (1980), in which idioms are stored as single wordlike units; *compositional* models, in which analysis of an idiom's individual words is necessary to comprehend the idiom's figurative meaning (Gibbs, Nayak, & Cutting, 1989); and *hybrid* models, in which a compositional analysis of an idiom's words and the retrieval of the idiom's figurative meaning happen simultaneously, such as Cacciari and Tabossi's Configuration Hypothesis (1988) or Titone and Connine's Hybrid Model (1999). These types of models differ significantly. However, they all acknowledge the tension between an idiom's overall figurative meaning and the meanings of its individual words, and identify this tension as a difficulty that any model of idiom representation must explain. Accounting for this tension has resulted in most models of idiom representation being isolated from models of language representation in general.

However, recent research in several areas of literal language processing has brought literal and figurative language research closer together. One such area is research into ambiguous word processing. Homonyms are words like *bank*, which have multiple unrelated meanings. Polysemes are words like *sheet*, which have multiple related senses. However, these senses may be more or less literal. Studies of polysemy frequently acknowledge the difference between more literal polysemy—for example, *sheet* referring literally to both a sheet of paper and a bedsheet—and more figurative polysemy—for example, *eye* referring literally to a visual organ and metaphorically to a hole in a needle for thread (Frisson & Pickering, 1999; Klepousniotou, 2002;

Klepousniotou & Baum, 2007). Regardless of literality, these words are still all considered to be polysemous, and therefore more similar to each other than different.

To explore the effects of sense literality in polysemes, Klepousniotou (2002) conducted a cross-modal priming task comparing different kinds of polysemes to homonyms. In particular, she tested responses to metonymous polysemes such as *turkey* (the *animal*; metonymic extension of the *animal's meat*),- metaphorical polysemes such as *eye* (literal sense of *visual organ*; metaphorical extension of hole *in a needle*), and homonyms such as *pen* (*writing implement*; *enclosure*). She found significantly greater priming effects for metonymous polysemes compared to homonyms. However, priming effects for metaphorical polysemes were between those for metonyms and homonyms and were not significantly different from either. This suggests that literality in ambiguous words is a continuum rather than a strict division, with metonyms being the most figurative, homonyms the most literal, and metaphors occupying a flexible space between the two. This characterization of single-word ambiguity creates a precedent for consideration of literal and figurative language in the same sphere and as subject to the same processes, and invites comparison of other aspects of figurative language with potential analogues in literal processing.

A second point of comparison between literal and figurative language is research on literal multiword phrases: the same tensions between individual word meaning and overall phrase meaning that characterize idioms may also exist in literal language. Moreover, multiword units may have the same psychological salience as single words and may be equally important during language comprehension. Thinking of literal and figurative multiword phrases as more similar than different may help drive our understanding of how multiword phrases in general are processed.

The incorporation of metonyms and metaphorical polysemes into research on ambiguous word comprehension, as well as the similarities between literal and figurative multiword phrases, invites a characterization of idioms as extremely well-learned ambiguous multiword phrases. Characterizing idioms in this way allows specific predictions to be made about idiom processing: idioms and literal multiword phrases should behave similarly during comprehension, and factors influencing ambiguous word processing should influence idiom processing in similar ways.

Meaning Dominance and Meaning Relatedness

If ambiguity resolution proceeds similarly for both literal and figurative language, then the same constructs should produce similar effects on comprehension of literal and figurative units. Two constructs that have robust effects on comprehension of ambiguous words are meaning dominance and meaning relatedness. In this section, we examine whether analogous (and possibly identical) constructs affect comprehension of idioms, and, if so, whether their effects on idiom comprehension and on ambiguous words might be the same. Under a view of language processing in which the same mechanisms underlie literal and figurative ambiguity resolution, and multiword units are processed similarly to words, constructs affecting single-word comprehension should also affect multiword units. These effects should manifest regardless of whether the ambiguous unit is literal or figurative.

One construct that has robust effects on ambiguous word processing is the degree of semantic relatedness between the word's meanings or senses. In general, the high semantic relatedness between a polyseme's senses is thought to aid processing, resulting in easier processing of polysemes (for an overview, see Eddington & Tokowicz, 2015). In contrast, the low semantic relatedness between the meanings of a homonym results in no advantages or processing disadvantages compared to unambiguous words. For example, Klepousniotou and Baum (2007) found advantages only for polysemes, not for homonyms, compared to unambiguous words in both visual and auditory lexical decision. They interpreted this result as indicating that the separately-represented meanings of homonyms compete for activation when a homonym is encountered.

Although previous research on idiom comprehension has not identified a single construct that is analogous to meaning relatedness in ambiguous words, there are several similar constructs that, when taken together, approximate meaning relatedness. These constructs are all used to explain how an overall figurative meaning is computed from the individual meanings of the idiom's words. One such construct is *transparency,* or how easily the comprehender can guess at the idiom's origin (Nunberg, Sag, & Wasow, 1994). A similar construct is *decomposability*, which is used either to measure how well individual words in the idiom metaphorically correspond to aspects of the idiom's figurative meaning (Gibbs et al., 1989; Nunberg et al., 1994), or to indicate more generally that an idiom's words contribute to the overall figurative meaning in some way (Caillies & Butcher, 2007; Hamblin & Gibbs, 1999; Titone & Connine, 1999). Idioms like *break the ice* or *sing the blues* are decomposable, and idioms like *kick the bucket* or *chew the fat* are generally characterized as nondecomposable (but see Nordmann, Clelland, & Bull, 2014, for a discussion of the difficulty inherent in decomposability classification). Neither transparency nor decomposability directly corresponds to meaning relatedness, but both are concerned with the semantic relationship between literal and figurative meanings. Both may therefore be considered proxies for meaning relatedness: examination of the effects of transparency and decomposability on idiom processing can inform understanding of how idioms are processed and represented.

The effects of decomposability in idioms are strikingly similar to the effects of meaning relatedness in ambiguous words. Several studies find that decomposable idioms are comprehended more quickly than nondecomposable idioms (Caillies & Butcher, 2007; Gibbs et al., 1989). To explain this phenomenon, Titone and Connine (1999) suggested that the literal and figurative meanings of decomposable idioms were highly semantically related, and that this relatedness sped comprehension of decomposable idioms. They proposed the Hybrid Model (1999), in which idiom comprehension takes two simultaneous routes: direct access of the idiom's meaning, and compositional analysis of the idiom's individual words. Under their view, slower processing of nondecomposable idioms is caused by interference between the directly-retrieved figurative meaning and the highly semantically dissimilar literal meaning, which is activated concurrently during processing. In contrast, they propose that meanings of a decomposable idiom are highly similar, and concurrent compositional analysis of the literal meaning augments direct retrieval of the figurative meaning, resulting in faster processing.

Comparison of studies of idiom decomposability and meaning relatedness in ambiguous words reveal striking processing similarities between these two types of ambiguity. In particular Titone and Connine's (1999) test of their Hybrid Model and Brocher, Foraker, & Koenig's (2016) examination of homonyms and polysemes comprehension in reading find similar patterns of results using broadly similar study designs. Titone and Connine (1999) examined reading times for decomposable and nondecomposable idioms. Idioms were presented accompanied by a context sentence; this sentence appeared either before or after the idiom, and biased either the literal or figurative meaning. Titone and Connine (1999) found that nondecomposable idioms were read more slowly when context preceded the idiom, regardless of contextual bias. However, decomposable idioms were read equally quickly regardless of both contextual bias and location of the context. They interpreted these results as suggesting that both literal and figurative meanings of the idiom were activated during comprehension. This resulted in no processing costs for decomposable idioms because of high degree of relatedness between their meanings. However, integration of the contextually-appropriate meaning of a nondecomposable idiom was impaired because of competition between the unrelated meanings, resulting in slower reading times.

Brocher and colleagues (2016) examined reading times for homonyms and polysemes embedded within sentences. Critically, these sentences contained disambiguating regions that appeared either before or after the ambiguous word. Homonyms showed longer reading times compared to their unambiguous control words regardless of the location of the disambiguating region, similar to the slow-down for nondecomposable idioms found by Titone and Connine when context was presented before the idiom (1999). Polysemes, however, showed overall less difficulty, similar to the easy processing of decomposable idioms found by Titone and Connine. Brocher and colleagues interpreted these results as demonstrating facilitated processing for the semantically related senses of a polyseme. Although there are differences in the designs of these two studies, most particularly in the locations of the disambiguating regions, the correspondences in design and results are compelling enough to predict further correspondences in future research. These correspondences, if they exist, would support a model of language comprehension in which the same mechanisms underlie both literal and figurative ambiguity resolution, at both the single word and multiword levels.

A second construct that affects processing of ambiguous words, and may have parallel effects on idiom processing, is meaning dominance: one meaning of an ambiguous word is often dominant over another, and meaning dominance interacts with semantic relatedness during comprehension of ambiguous words. Examining the effects of both meaning dominance and meaning relatedness on ambiguous word comprehension is often more informative about how these words are processed than examination of one factor alone. Foraker and Murphy (2012) embedded ambiguous words in contexts that supported either the word's dominant meaning (for example, the fabric meaning of *cotton*) or subordinate meaning (the crop meaning of *cotton*), or in neutral contexts. They found speeded processing, as indexed by reading times and eye movement patterns, when the context supported the word's dominant meaning compared to the subordinate meaning. Critically, they also found that sense similarity interacted with dominance to affect several eyetracking measures of early processing: words with highly related senses, but with one sense strongly dominant over the other (eg. *gem*[1]), showed a processing disadvantage. Foraker and Murphy explained this by proposing that there is more competition between senses when one sense is very dominant but sense similarity is also very strong.

Interestingly, studies of word ambiguity overwhelmingly find processing advantages for polysemous words and either disadvantages or no effects for homonymous words (for a review, see Eddington & Tokowicz, 2015). The disadvantage for polysemes with one highly dominant sense found by Foraker and Murphy is unusual compared to the polyseme advantage usually seen in the word ambiguity literature, and seems more similar to the processing disadvantage for homonyms compared to polysemes found in other studies. In short, polysemes in general may be advantaged over homonyms during processing, but a subset of polysemes with one highly dominant sense occupy a middle ground in which their effects on processing are more akin to

---

[1] This example is taken from the stimuli of the present study because Foraker and Murphy did not include their stimuli in their article.

homonyms. This is further evidence that the division between homonyms and polysemes may be continuous rather than categorical, and that meaning dominance and sense similarity likely interact during comprehension to drive sense selection.

Although the research on meaning dominance in idioms is less comprehensive than corresponding research in ambiguous words, there are indications that idioms may have one meaning that is more dominant over the other, and processing of the idiom may differ depending on which meaning is biased by the context. Idioms are interpretable as figurative phrases even when they appear in isolation, without contextual support biasing the comprehender towards a figurative meaning. This indicates that an idiom's overall figurative meaning can be dominant over its compositional literal meaning. This "figurative-dominant" perspective is further supported by the observation that knowing an idiom's figurative interpretation appears to suppress comprehenders' ability to recognize that the idiom also has a literal interpretation (Gibbs, 1980). This view is reflected in older models of idiom representation, such as Swinney and Cutler's Lexical Representation Hypothesis (1979). Under this model, the figurative meanings of idioms are stored as large, lexicalized "chunks", akin to long words, resulting in a dominant figurative meaning. Because accessing a single lexical entry is faster than compositionally analyzing literal meanings of multiple words, idioms are processed faster than literal strings. This is congruent with literature that finds rapid access of idiomatic meaning even in isolation: supportive context is not necessary for an idiomatic interpretation if phrases and their figurative meanings are stored in the lexicon. However, the context in which the idiom is presented may sometimes be strong enough to override the idiom's dominant figurative bias. For example, Holsinger (2013) found that participants looked at figurative probes when they heard idioms embedded in figurative contexts, and at literal probes when they heard idioms embedded in literal contexts. However, he did not quantify whether the idioms in question were truly figurative-dominant, or if their meanings were more balanced.

In conclusion, there are hints that the constructs of meaning dominance and meaning relatedness may be common to both ambiguous words and to idioms, and may have similar effects on ambiguity resolution regardless of whether the ambiguous unit is literal or figurative, single word or multiword. Foraker and Murphy (2012) found that meaning relatedness and meaning dominance interacted during comprehension of ambiguous words, suggesting that, as one meaning becomes more dominant, polysemes are processed more similarly to homonyms. Finding the same pattern during idiom comprehension would be strongly suggestive evidence for a flexible language system that resolves ambiguities similarly regardless of their literality or grain size. This would also point to a characterization of figurative language as most similar to literal language than different. To investigate this hypothesis, we conducted two experiments.

Experiment 1 examined potential parallels between the processing of ambiguous words and the processing of idioms. In this experiment, we looked for similar patterns of meaning activation during the priming of idioms and ambiguous words. To do this, we conducted two parallel experiments. **Experiment 1A** used a word-to-word priming paradigm and lexical decision task to examine effects of meaning relatedness and meaning dominance on facilitation of ambiguous word meanings. **Experiment 1B** used an analogous phrase-to-word priming paradigm and lexical decision task to investigate how the same constructs influence facilitation of idiom meanings. Although idioms and ambiguous words are different enough that designing an experiment to directly compare them would be prohibitively difficult, these parallel experiments allows a close comparison between idiom processing and single word processing. Similar effects of meaning dominance and meaning relatedness on processing idioms and ambiguous words would suggest that these two types of ambiguous units are being treated similarly by the language processing system. Different patterns of priming would suggest that these types of literal and figurative language are distinct from one another, and are treated differently by the language processing system.

Experiment 2 explored the way meaning relatedness and meaning dominance influence idiom comprehension by examining patterns of eye movements during idiom reading. We used a design incorporating elements of Foraker and Murphy's (2012) and Brocher and colleagues' (2016) studies investigating eye movements in response to ambiguous words in context. Following Foraker and Murphy's findings, we expect that idioms with one highly dominant meaning and overall highly semantically related meanings will elicit greater disruption to early eye movement measures than will idioms with less dominant meanings. Finding particularly this pattern of disruption would be suggestive evidence that the same constructs influence ambiguity resolution regardless of literality. Experiment 2 will also allow us to investigate the time course along which meaning relatedness and meaning dominance affect idiom comprehension. If more dominant meanings of idioms are more lexicalized and therefore accessed more quickly (Gibbs, 1980), then dominance effects might emerge in earlier eye movement measures. In contrast, if meaning relatedness can only be computed post-phrase (Libben & Titone, 2008), then meaning relatedness effects might only be seen in later eye movement measures.

2.0 EXPERIMENT 1

Experiment 1 examines potential parallels between the processing of ambiguous words and the processing of idioms using two parallel primed lexical decision experiments. In both experiments, we use ambiguous units (idioms and ambiguous words) as primes and look for processing facilitation, as influenced by meaning dominance and meaning relatedness, on target words related to the different meanings of each. We also compare processing facilitation following ambiguous units to unambiguous control units. This design enables us to compare processing of ambiguous words and idioms as directly as possible.

Similar effects of meaning dominance and meaning relatedness following idiom and ambiguous word primes would indicate that these constructs have comparable effects on processing of ambiguous units, and therefore that these types of literal and figurative language are treated the same in ambiguity resolution. Different patterns of facilitation would suggest that these types of literal and figurative language are distinct from one another, and are treated differently by the language processing system.

Previous research has shown that meaning dominance and meaning relatedness (usually indexed based on whether a word is a homonym or polyseme) interact during ambiguous word processing. In particular, having one strongly dominant meaning seems to have a greater impact on processing for homonyms than for polysemes. For example, Frazier and Rayner (1990) found easier processing for the dominant meanings compared to the subordinate meanings of homonyms; in contrast, the two senses of polysemes were equally easily comprehended. Similarly, Klepousniotou and colleagues (2008) found greater effects of dominance for homonyms compared to polysemes. Brocher, Foraker, and Koenig (2016) also investigated meaning dominance and meaning relatedness in their study of irregular polysemes, although they compared only neutral and subordinately-biased context sentences. They found greater processing difficulty after subordinately-biased contexts for homonyms compared to polysemes, suggesting that the greater relatedness between polyseme senses aided comprehension even when the subordinate sense was biased. Although meaning dominance may have greater effects on homonym processing, evidence exists showing that meaning dominance can affect polyseme processing as well. Foraker and Murphy (2012) embedded polysemes into neutral and biased sentence contexts and found overall effects of dominance such that dominant polyseme senses were accessed more easily, even after neutral contexts. However, they also found that dominance interacted with sense similarity: polysemes with one highly dominant sense but high sense similarity were the most difficult to interpret. Taken together, these studies indicate that meaning dominance matters more for processing when meanings are less related. However, meaning dominance can still affect processing of ambiguous words with more related meanings if one meaning is strongly dominant.

Meaning dominance and meaning relatedness may affect processing of idioms and ambiguous words in similar ways. In particular, many studies have found advantages for decomposable idioms, or idioms that have strong or easily-recognizable relationships between their literal and figurative meanings (Caillies & Butcher, 2007; Gibbs, Nayak, & Cutting, 1989; Titone & Connine, 1999)—similar to the advantage for polysemous words over homonymous words found in lexical decision tasks, possibly due to the greater semantic relatedness between their senses (for a review, see Eddington & Tokowicz, 2015). Additionally, although little research has directly investigated the effects of meaning dominance on idiom processing, many models of idiom comprehension have either implicitly assumed that an idiom's figurative meaning is the dominant meaning—for example, Swinney and Cutler's Lexical Representation Hypothesis (1979)—or proposed that the degree to which an idiom is identifiable as an idiom—a measure called conventionality, and a reasonable proxy for dominance—directly influences comprehension (Titone & Connine, 1999). Again, however, little research has been done investigating how meaning relatedness and dominance of the figurative meaning work together to facilitate or inhibit idiom comprehension. In the present study, we therefore look to the ambiguous word literature to make predictions about the interactive effects of meaning dominance and meaning relatedness on idiom comprehension.

The present study consists of two parallel primed lexical decision experiments. Each experiment uses ambiguous units as primes and compares responses to two target words. In Experiment 1A, we use ambiguous words as primes and target words related to the dominant and subordinate meanings of the prime word. In Experiment 1B, we use idioms as primes, and target words are related to the literal and figurative meanings of the idioms. Although this design does not allow us to measure responses to the ambiguous units themselves, it does enable us to determine whether previous exposure to an ambiguous unit facilitates processing for either or both target units. We also use unambiguous units—either single words or multiword phrases—as control primes. This enables us to compare responses to each target when primed by either ambiguous or control units, and therefore to determine both whether ambiguous units have an advantage in priming, and if this effect is different for different targets. Presentation of the ambiguous prime should activate its different meanings to different degrees depending on their dominance and relatedness. When there is greater overlap between the target word and the activated meaning of the ambiguous prime, greater priming should result.

Additionally, previous research has shown different effects of meaning relatedness depending on the amount of semantic engagement required by the task (Armstrong & Plaut, 2016). We therefore manipulate semantic engagement by changing the average bigram frequency of the nonword targets in filler trials. This manipulation allows us to measure responses at two discrete points on the processing time continuum while still using a single ISI, making it more likely that we will see effects of both polysemy and homonymy. Armstrong and Plaut showed that participants process more deeply when nonwords have higher bigram frequencies and thus harder to distinguish from real words, thereby forcing participants to rely on semantics rather than surface features to make lexical decisions. This forced deep processing results in disadvantages for homonymous words compared to unambiguous words because the two largely unrelated meanings of the homonym must be active and compete. In contrast, nonwords with lower bigram frequencies are easier to identify as nonwords and therefore only shallowly engage semantic processing. This results in an advantage for polysemous words compared to unambiguous words because there is little competition between their very similar senses.

Overall, we expect to see similar patterns of results for ambiguous units regardless of grain size (single word vs. multiword) and literality (literal vs. figurative). First, we predict that participants will respond more quickly and more accurately to target words following ambiguous primes compared to control primes because the target words are related in meaning to the ambiguous primes but not the control primes. We also predict an interaction between prime type and target type such that the greatest facilitation will occur after ambiguous primes for the dominant-related (in ambiguous words) or figurative-related (in idioms) target words. This is because when an ambiguous unit is presented in isolation, as in the priming paradigm, the dominant/figurative meaning is more strongly activated than the subordinate meaning (Brocher et al., 2016; Gibbs, 1980; Klepousniotou et al., 2008; Titone & Connine, 1999).

We also predict that greater relatedness between meanings will facilitate priming for both dominant- and subordinate-related target words in ambiguous words (Eddington & Tokowicz, 2015) and idioms (Caillies, & Butcher, 2007; Gibbs, Nayak, & Cutting, 1989; Titone & Connine, 1999), because said greater relatedness will induce less competition between meanings. We also predict several interactions. First, we predict an interaction between target type (dominant/figurative vs. subordinate/literal) and meaning relatedness: when an ambiguous unit's meanings are highly related, priming should be similar for both dominant- and subordinate-related targets. In contrast, when meanings are highly unrelated we expect to see slower reaction times for subordinate-related targets (Klepousniotou, Pike, Steinhauer, & Gracco, 2012).

We further predict an interaction between target type and meaning dominance: target type will matter more for ambiguous units with one strongly dominant meaning than for ambiguous units with more balanced meanings. Specifically, when meanings are balanced, one meaning is not strongly dominant over the other, and therefore even the subordinate meaning is still easily activated. Thus, both the dominant- and subordinate-related targets should be processed with similar ease. When meanings are strongly biased, however, we expect to see slower RTs following subordinate-biased targets (Eddington & Tokowicz, 2015).

In sum, finding these parallel effects in both idioms and ambiguous words would be suggestive evidence that the language processing system treats these two types of ambiguous units similarly during ambiguity resolution.A summary of Experiment 1A and 1B results can be viewed in Table 1.

Table 1. Experiment 1A and 1B result summary

Dependent Measure	Test Type	Ex 1A Result Summary	Ex 1B Result Summary
Reaction Time	Priming	Faster RTs after ambiguous words than control words	Faster RTs after idioms than control phrases
	Meaning Dominance & Meaning Relatedness	**Homonyms:** faster RTs with increased dominance **Polysemes:** *Higher Nonword Freq:* Faster RTs with increased dominance *Lower Nonword Freq:* No effects	No effects
Accuracy	Priming	Fewer errors after ambiguous words than control words	**Higher Nonword Freq:** Fewer errors after idioms than control phrases **Lower Nonword Freq:** No effects
	Meaning Dominance & Meaning Relatedness	**Homonyms:** higher accuracy with increased dominance **Polysemes:** higher accuracy when nonwords were higher frequency	Most errors when figurative dominance was high and responding to literal targets

2.1 EXPERIMENT 1A

Methods

2.1.1.1 Materials

Participants completed a lexical decision task using semantic priming. Sample critical and filler items can be viewed in Table 2. Prime words were 72 ambiguous words taken from

Eddington (2015). Each ambiguous prime word was paired with an unambiguous control word. Ambiguous primes and unambiguous controls were matched on length, concreteness (Brysbaert, Warriner, & Kuperman, 2014), and frequency (SubtlexUS; Brysbaert & New, 2009). Each ambiguous prime word was paired with two target words, one related to the ambiguous word's dominant meaning, and one related to the subordinate meaning (this manipulation is hereafter referred to as the "target type" manipulation). Stimulus presentation was counterbalanced such that each participant only saw one of the possible four prime-target pairs. Descriptive statistics and norm values for all items can be viewed in Appendix B.

We used real word and nonword filler items. All participants saw all word fillers. Lists were counterbalanced across participants such that each participant either saw only high bigram or low bigram nonwords; this manipulated semantic engagement between participants.

Real-word fillers were 36 unambiguous word primes. Half were paired with a target that was related to the meaning of the prime, and half were paired with an unrelated target. Primes and targets were roughly matched on frequency and concreteness.

Nonword fillers were 108 real word primes, each paired with two possible nonword targets. Nonwords were created using Wuggy (Keuleers & Brysbaert, 2010) a nonword generator. We created a set of 108 real words to use as primes, and then a second set of real words that were matched to them on length, concreteness (Brysbaert et al., 2014), and frequency (Brysbaert & New, 2009). These words were used as input values to generate nonwords in Wuggy. We generated 50 possible nonwords for each input real word. For each set of nonwords, we identified the two nonwords with the highest and lowest mean bigram frequency using the English Lexicon Project (Balota et al., 2007), excluding those nonwords that were homophonous with a real English word (Armstrong & Plaut, 2016), as well as false plurals and gerunds. We created two lists of nonwords: high bigram frequency (M = 2676.95; SD = 736.38) and low bigram frequency (M = 860.68; SD = 506.59). The two groups of nonwords significantly differed in mean bigram frequency ($t(107) = -29.08$; $p<.05$).

Table 2. Sample stimuli for each condition (Experiment 1A)

Manipulation	Item Type	Prime Type	Target Type	# of Stimuli	Example (prime:TARGET)
Within-subjects	Critical	Ambiguous Word	Dominant-related	18	staff : TEACHER
			Subordinate-related	18	staff : STICK
		Control Word	Dominant-related	18	punch : TEACHER
			Subordinate-related	18	punch : STICK
	Filler	Real Word	Related Word	18	roof : FLOOR
			Unrelated Word	18	spine : KNITTING
Between-subjects			High bigram frequency nonword	108	shiny : SMERSED
			Low bigram frequency nonword	108	shiny : SMELFTH

2.1.1.2 Norming

Meaning Dominance

To determine which meaning or sense of the ambiguous words was the dominant one, we conducted a norming study. 25 undergraduate students from the University of Pittsburgh participated for course credit using a Qualtrics survey. Participants were shown a word and told that it could have multiple meanings. They then moved sliders indicating what percentage of the time they expected the word to have each of two meanings. For example, participants might view the word "cotton" and be asked what percentage of the time they expected "cotton" to mean "a fiber used to make clothing" versus "the plant that produces those fibers". Sliders could not be moved to equal more than 100%, and the participants could not move to the next question if the slider values equaled less than 100%. Meaning dominance norm values were used as continuous predictors in later statistical analyses. Descriptive statistics for dominant and subordinate meanings can be viewed in Table 3.

Table 3. Descriptive statistics for meaning dominance norm

Meaning	Mean	SD	Range
Dominant	54.5	3.17	50.04 - 63.04
Subordinate	45.5	3.17	36.96 - 49.96

Prime-Target Relatedness

The same 25 participants who completed the meaning dominance rating survey also completed a survey of prime-target relatedness. An additional 25 participants also completed the prime-target relatedness survey. Participants were shown pairs of words and asked to rate how related the words were on a scale of 1 to 7 (1 = *very related*, 7 = *very unrelated*). Five pairs of relatedness comparisons were collected. We collected relatedness comparison ratings between the ambiguous prime words and their dominant- and subordinate-related target words, as well as relatedness comparison ratings between the unambiguous control prime words and the dominant-related and subordinate-related target words. Additionally, we collected relatedness comparison ratings between the ambiguous prime words and their matched unambiguous control prime words. Prime-target relatedness norm values were used as continuous predictors in later statistical analyses. Descriptive statistics for all comparisons can be viewed in Table 4.

Table 4. Descriptive statistics for prime-target relatedness norm comparisons

Comparison	Mean	SD	Range
Ambiguous/ Dominant	1.99	.67	1 - 5
Ambiguous/ Subordinate	2.31	.81	1 - 4.5
Ambiguous/ Control	5.58	.86	2.8 - 7
Control/ Dominant	5.60	.76	3 - 6.9
Control/ Subordinate	5.7	.75	4 - 7

2.1.1.3 Procedure

Forty-nine undergraduate students from the University of Pittsburgh who had not participated in the norming completed the experiment for course credit. Before participating in the study, all participants provided informed consent and completed a questionnaire collecting demographic information such as age and language background. All participants were native speakers of English.

Participants viewed items on a personal laptop computer using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA), and responded using the “1” and “5” keys on a button box (Psychology Software Tools, Pittsburgh, PA). Following consenting and collection of demographic information, participants completed the priming task. Experiments 1A and 1B were completed sequentially, and their order was counterbalanced across participants. Instructions were displayed on the screen and read aloud to participants, followed by ten practice items.

Following completion of the practice items, the experimenter verified that participants understood the study procedures before starting the main experiment. Participants saw 216 randomly ordered prime-target pairs of the types described above. Each trial began with a blank screen displayed for 250 ms, followed by a centrally-located fixation cross displayed for between 750 and 950 ms (see Armstrong & Plaut, 2016). Following a 100 ms inter-stimulus interval (ISI), prime words were displayed for 250 ms, followed by a 200 ms ISI. Targets were displayed until a lexical decision was made or 4000 ms had passed. Participants completed the experiment in one sitting, but were encouraged to take a break between Experiments 1A and 1B.

Results

Data were analyzed using linear mixed effects models (reaction time) and generalized linear mixed effects models (accuracy; Baayen, 2008) in the R statistical computing package (R Development Core Team, 2013; ver. 3.0.1) and using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015; ver. 1.1-7). P-values were obtained using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2016; ver. 2.0-20). Models were fit using the fullest random effects structure that would support convergence (Barr, Levy, Scheepers, & Tily, 2013). Models contained fixed effects of prime word meaning dominance, prime word ambiguity type (homonym vs. polyseme), target type (dominant-related vs. subordinate-related), nonword filler type (high vs. low mean bigram frequency), trial number, prime-target relatedness norm score, previous trial reaction time, previous trial accuracy, target word bigram frequency, target word length, number of syllables in the target word, and target word concreteness. Models additionally contained random intercepts of participant and item. Finally, we included random slopes of trial number, meaning dominance, ambiguity type, prime-target relatedness norm score, nonword bigram frequency, and target type within participants, and random slopes of target type within items. In cases of non-convergence, the random slopes that explained the least amount of variance were removed until convergence was achieved. Outcome variables were reaction time to make a lexical decision and accuracy of lexical decision. All fixed effects except for trial number and previous trial accuracy were z-score transformed to aid convergence. Finally, we used the inverse of the reaction time outcome variable and the fixed effects of previous trial reaction time and prime-target relatedness norm score, again to aid convergence and increase interpretability.

2.1.1.4 Reaction Time

Ambiguous Word Priming Effects

The first analysis examined the effects of ambiguity on reaction time. We removed filler trials (7,056 trials) and incorrect trials (153 trials). We also removed trials that had reaction times greater than 2.5 standard deviations outside the means by participant and list (18 trials). Overall, 4.8% of trials were removed during trimming. After trimming, 3,357 total trials were analyzed.

We created four comparisons of interest using contrast coding. We compared RTs following the unambiguous control prime words to those following the ambiguous prime words. We predict that participants will respond more quickly after seeing ambiguous prime words, regardless of target type. We compared RTs in response to a subordinate meaning-related target to those in response to a dominant meaning-related target. This comparison should interact with prime type: when prime words are ambiguous, participants will respond faster to dominant-related targets compared to subordinate-related targets. The third comparison compared RTs in the higher nonword frequency lists to the lower nonword frequency lists, using higher frequency nonwords as the baseline. This enabled us to determine whether the bigram frequency of the nonword fillers affected how participants responded to target words with different characteristics. Finally, the fourth comparison compared RTs when the previous trial's accuracy was correct to when the previous trial's accuracy was incorrect, using incorrect trials as the baseline; this comparison was included as a control. The interaction of target type (dominant-related vs. subordinate-related) and nonword bigram frequency was included in the models.

In the analysis of ambiguity of reaction time, we did not include the fixed effects of meaning dominance and ambiguity type because control trials did not have associated dominance and ambiguity type values. We also included the fixed effect of prime type (ambiguous vs. unambiguous control). The model did not contain any random slopes due to convergence issues. Descriptive statistics for reaction time by prime type and target type can be viewed in Table 5.

Table 5. Descriptive statistics for Ex1A ambiguous and control RTs (ms)

Prime Type	Target Type	Mean	SD
Ambiguous	Dominant	588	182.01
	Subordinate	598	246.73
Control	Dominant	596	187.42
	Subordinate	608	201.66

Model results can be viewed in Appendix A. As predicted, there was a significant effect of prime type: participants were faster to respond after ambiguous words than they were to the control words ($\hat{\beta}$=-.07; SE=.02; t=-3.15; p<.05). There were no other significant effects, and no interactions. This represents a classic priming effect, given that the targets were related in meaning to the ambiguous words but not the control words.

Meaning Dominance and Ambiguity Type (Homonyms vs. Polysemes)

The second analysis investigated effects of meaning dominance and ambiguity type on reaction time (RT). Control trials involved unambiguous words that did not have associated dominance and ambiguity type values, and were therefore removed from analysis. We also removed filler trials and incorrect trials. Finally, we removed trials that had reaction times greater than 2.5 standard deviations outside the means by participant and list; this removed 18 trials from analysis. After trimming, 1,695 total trials were analyzed. Descriptive statistics for reaction time by meaning dominance, ambiguity type, and target type can be viewed in Table 6 Although meaning dominance was treated continuously in analyses, for ease of presentation it appears in Table 6 using a median split.

Table 6. Descriptive statistics for Ex1A reaction times (ms)

Target Type	Meaning Dominance	Ambiguity Type	Mean	SD
Dominant	High	Polyseme	578	180.57
		Homonym	569	170.11
	Low	Polyseme	605	195.22
		Homonym	614	184.14
Subordinate	High	Polyseme	584	180.62
		Homonym	601	214.59
	Low	Polyseme	614	292.76
		Homonym	593	260.42

The dependent measure was reaction time. The model contained the random slopes of meaning dominance, ambiguity type, prime-target relatedness score, and nonword bigram frequency within participants; more complete models did not converge. Model estimates for all fixed effects and interactions can be viewed in Appendix A. Significant effects are bolded; marginal effects are italicized. There was a significant effect of meaning dominance such that as dominance increased, reaction time decreased ($\hat{\beta}$=-.01; SE=.04; t=-1.94; p<.05). There was also a significant three-way interaction between meaning dominance, ambiguity type, and nonword bigram frequency ($\hat{\beta}$=.04; SE=.02; t=1.98; p<.05). Finally, there were two marginally significant three-way interactions: one between meaning dominance, ambiguity type, and target type ($\hat{\beta}$=-.05; SE=.03; t=-1.79; p=.08), and one between meaning dominance, target type, and nonword bigram frequency ($\hat{\beta}$=.04; SE=.02; t=1.83; p=.07). There were also several significant effects of control variables.

To investigate the interactions, we separated the data by ambiguity type (homonyms vs. polysemes). Model estimates for all fixed effects and interactions within each ambiguity type can be viewed in Appendix A.

For the model examining words with highly-related meanings (polysemes), we included fixed effects of meaning dominance, nonword filler type (high vs. low mean bigram frequency), prime-target relatedness score, trial number, previous trial reaction time, previous trial accuracy, target word bigram frequency, target word length, number of syllables in the target word, and target word concreteness, and random intercepts of participant and item. Models containing random slopes did not converge. There was a significant interaction between meaning dominance and nonword bigram frequency ($\hat{\beta}$=.03; SE=.02; t=2.13; p<.05). This interaction can be visualized in Figure 1.

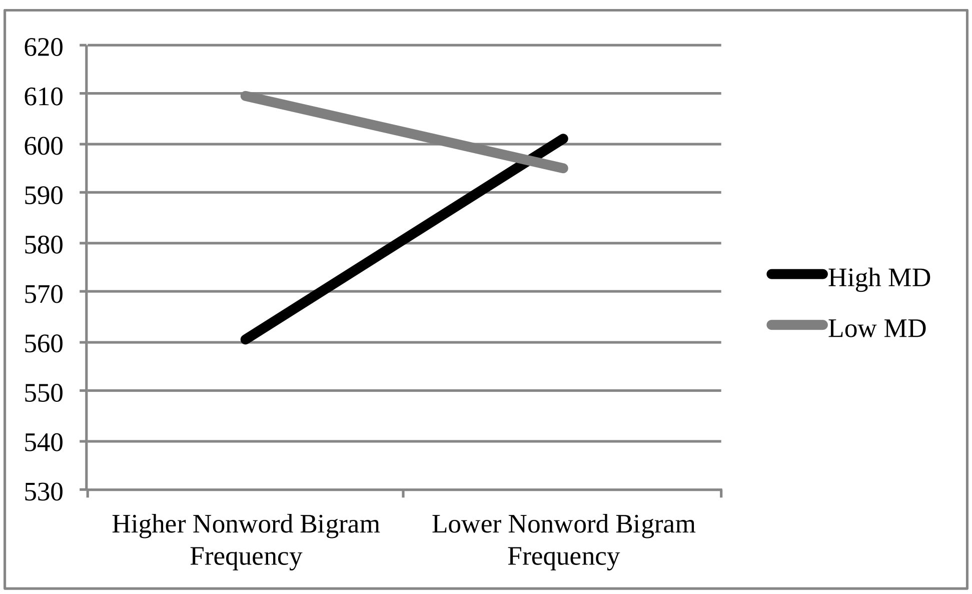

Figure 1. Effects of Meaning Dominance and Nonword Bigram Frequency on RT within Polysemes

To investigate this interaction, we separated the data by nonword bigram frequency. For the model looking at effects of meaning dominance within the higher bigram frequency nonword condition, we included the fixed effects named above, with the exception of nonword filler type. Models containing random slopes did not converge. There was a marginal effect of meaning dominance such that as dominance increased, reaction time decreased ($\hat{\beta}$=-.01; SE=.01; t=-1.97; $p<.05$). This result is unexpected: effects of meaning dominance appear more commonly in homonyms than polysemes, and strong meaning dominance has been shown to slow processing of polysemes in reading (Foraker & Murphy, 2012). However, the fact that this effect appears only when nonwords are higher frequency suggests that semantic engagement may play a role in explaining this effect. We will return to this point in the Discussion for Experiment 1A.

For the model looking at effects of meaning dominance within the lower bigram frequency nonword condition, models containing random slopes likewise did not converge. There were no significant effects.

For the model examining words with less related meanings (homonyms), we included the fixed effects of meaning dominance, nonword filler type (high vs. low mean bigram frequency), prime-target relatedness score, trial number, previous trial reaction time, previous trial accuracy, target word bigram frequency, target word length, number of syllables in the target word, and target word concreteness, and random intercepts of participant and item. Models containing random slopes did not converge. There was a significant effect of meaning dominance such that as dominance increased, reaction time decreased ($\hat{\beta}$=-.01; SE=.01; t=-1.97; $p<.05$). This result is congruent with other research finding advantages for the dominant meanings of homonyms. There were no other effects of study variables, and no interactions.

2.1.1.5 Accuracy

Accuracy results grouped by prime type and target type can be viewed in Table 7.

Table 7. Experiment 1A accuracy results

Prime Type	Target Type	Proportion Correct
Ambiguous	Dominant	.96
	Subordinate	.97
Control	Dominant	.93
	Subordinate	.94

Ambiguous Word Priming Effects

The first analysis examined the effects of ambiguity on accuracy. Data were transformed using the empirical logit collapsed over participants to allow us to include item-level variables in the model. The model included the fixed effects described in the priming analyses above.

As predicted, there was a significant effect of prime type such that participants made fewer errors after seeing ambiguous words than they did after seeing the control words ($\hat{\beta}=.24$; $SE=.08$; $t=3.16$; $p<.05$). Although there were the expected effects of control variables, there were no other significant effects for study variables, and no interactions. Model results can be viewed in Appendix A.

Meaning Dominance and Ambiguity Type

The second analysis investigated effects of meaning dominance and ambiguity type on accuracy. Control trials involved unambiguous words that did not have associated dominance values and were therefore removed from analysis. Data were trimmed as described in the reaction time analyses above.

As in the priming analyses, these analyses were conducted over empirical logit-transformed data. The models contained fixed effects of meaning dominance, ambiguity type (homonym vs. polyseme), nonword bigram frequency (high vs. low), target type (dominant vs. subordinate), target word length, target word number of syllables, and target word concreteness.

Model results can be viewed in Appendix A. There was a marginally significant interaction between nonword bigram frequency and ambiguity type ($\hat{\beta}$=-.26; SE=.15; t=-1.71; p=.09). There was also a marginally significant interaction between target type and ambiguity type ($\hat{\beta}$=-.31; SE=.16; t=-1.94; p=.053). Finally, there was a marginally significant interaction between ambiguity type and meaning dominance ($\hat{\beta}$=-.15; SE=.08; t=-1.86; p=.06). Although there were some significant effects of control factors, there were no other significant main effects of study factors, and no interactions.

To investigate the marginally significant interactions, we looked for the effect of meaning dominance within each ambiguity type. We chose this variable because ambiguity type was involved in all the interactions. Additionally, we were most interested in differences between homonyms and polysemes.

For the model assessing words with higher-related meanings (polysemes), there was a marginal effect of nonword bigram frequency such that participants were more accurate when nonwords were higher frequency than when nonwords were lower frequency ($\hat{\beta}$=-.18; SE=.10; t=-1.73; p=.09). There were no other effects, and no interactions.

For the model assessing words with less-related meanings (homonyms), there was a marginal effect of meaning dominance such that accuracy increased as dominance increased ($\hat{\beta}$=.11; SE=.06; t=1.87; p=.06). There were no other effects, and no interactions.

Discussion

In the current study, we found that meaning dominance and ambiguity type differently affected participants' reaction times and accuracy judgments during primed lexical decision. In reaction time, we found an effect of meaning dominance in homonyms such that as meaning dominance increased, reaction time decreased. This is congruent with previous research finding facilitation for dominant meanings of homonyms (Klepousniotou et al., 2008): when a homonym is presented in isolation, the dominant meaning is more strongly activated than the subordinate meaning, resulting in increased facilitation with increased dominance. Interestingly, we found facilitation for polysemes with highly dominant senses as well, but only when filler nonwords were higher frequency. This result is unexpected given that few studies find effects of dominance in polysemes; it's thought that their related meanings facilitate activation regardless of which meaning is ultimately selected. Foraker and Murphy (2012) did find effects of dominance in polysemes, but in the opposite direction: the dominant meanings of polysemes elicited more disruption during processing. However, the current effect was only marginally significant, and should therefore not be weighed too heavily.

Within the accuracy outcome measure, we again found that increased dominance led to increased processing facilitation for homonyms. For polysemes, we found increased facilitation only when semantic engagement was high, as induced by nonword fillers of higher frequencies. This is the opposite of what we would expect given the predictions made by Armstrong and Plaut

(2012): polysemes should be advantaged when nonwords are lower frequency because semantic processing is only shallowly engaged and there is little competition between their two related senses.

Finally, participants responded more quickly and more accurately after seeing ambiguous primes than they did after seeing unambiguous control primes, although there was no interaction with target type. This is expected given that target words were more related to ambiguous primes than to control primes, and demonstrates that our manipulation was successful.

Taken together, the results of Experiment 1A are as we expected for homonyms but not for polysemes. Although we did not find all the effects we anticipated, the lack of these effects does not point to a need to revise theories of ambiguous word processing, but rather may be due to our experimental design choices. We will return to this point in the General Discussion for Experiment 1.

2.2 EXPERIMENT 1B

Methods

2.2.1.1 Materials

Participants completed a lexical decision task using semantic priming, parallel to Experiment 1A. Condition counts for all stimuli and sample items can be viewed in Table 8. Prime phrases were 64 idiomatic phrases generated using idiom dictionaries (Ammer, 2013; Spears, 2007). All idiom stimuli had well-formed literal interpretations. Each idiomatic prime phrase was paired with a literal control prime phrase that was not related to either of the target words. Idiomatic phrases and their control prime phrases were matched on number of words and syntactic structure. Each idiomatic prime phrase was paired with two target words: one related to its literal interpretation, and one related to its figurative interpretation. Stimulus presentation was counterbalanced such that each participant only saw one of the possible four prime-target pairs. Descriptive statistics and norm values for all items can be viewed in Appendix B.

There were two types of fillers: those with real-word targets and those with nonword targets. Real-word target fillers were 36 phrasal primes, 24 literal and 12 idioms. We included idioms in the filler items to reduce the likelihood that participants would notice the experimental manipulation. All phrasal primes were paired with two possible real-word targets, one related to the prime and one unrelated. In the case of the idioms, the related targets were related to the idiom's figurative meaning. Lists were counterbalanced across participants such that each participant saw only one of the two possible target words.

Finally, nonword fillers were 108 phrasal primes, each paired with two possible nonword targets. Phrasal primes consisted of 36 idioms and 72 literal phrases. Nonword targets were created using the same procedure as in Experiment 1A. Lists were counterbalanced across participants such that each participant either saw only high bigram or low bigram nonwords. High frequency nonwords ($M = 2683.14$; $SD = 703.23$) were significantly higher in mean frequency than low frequency nonwords ($M = 1222.11$; $SD = 557.95$; $t(107) = -25.35$; $p<.05$).

Table 8. Sample stimuli for each condition (Experiment 1B)

Manipulation	Item Type	Prime Type	Target Type	# of Stimuli	Example (prime:TARGET)
Within-subjects	Critical	Idiom	Figurative	18	sing the blues : DEPRESSION
			Literal	18	sing the blues : MICROPHONE
		Control Literal Phrase	Figurative	18	walk the dogs : DEPRESSION
			Literal	18	walk the dogs : MICROPHONE
	Filler	Phrase	Related Word	18	open the door : HINGE
			Unrelated Word	18	open the door : GNOME
Between-subjects			High bigram frequency nonword	108	smash the bug : SUKIMUM
			Low bigram frequency nonword	108	smash the bug : MAWISEM

2.2.1.2 Norming

Meaning Dominance

To determine the degree of dominance of each idiom's figurative meaning, we conducted a norming study. 57 undergraduate students from the University of Pittsburgh participated for course credit using a Qualtrics survey. 140 idioms were normed. Participants were shown an idiom and told that it could have multiple meanings. They then moved sliders indicating what percentage of the time they expected the idiom to have each of two meanings. For example, participants might view the idiom "have cold feet" and be asked what percentage of the time they expected "have cold feet" to mean "retreat from an undertaking" versus "your feet are cold". Sliders could not be moved to equal more than 100%, and the participants could not move to the next question if the slider values equaled less than 100%. Meaning dominance norm values were used as continuous predictors in later statistical analyses. Descriptive statistics for literal and figurative meanings can be viewed in Table 10.

Meaning Relatedness

The same 57 participants who completed the meaning dominance rating survey also completed a survey of meaning relatedness. Participants were shown an idiom's literal and figurative meanings and asked to rate how related they were on a scale of 1 to 7 (1 = *very related*, 7 = *very unrelated*). Participants were not told that the phrases they saw were idiom meanings. Meaning relatedness norm values were used as continuous predictors in later statistical analyses. Example idioms with high and low figurative dominance and meaning relatedness values can be viewed in Table 9. Descriptive statistics for the meaning relatedness norm can be viewed in Table 10.

Table 9. Example idioms with high and low dominance/relatedness values

Meaning Relatedness	Figurative Dominance	Idiom
High	High	on the fence
	Low	play with fire
Low	High	fall off the wagon
	Low	cut down to size

Idiom Familiarity

53 undergraduate students from the University of Pittsburgh participated for course credit using a Qualtrics survey. Participants rated the familiarity of each idiom on a scale of 1 (very familiar) to 7 (very unfamiliar). 240 idioms were normed. Idiom familiarity norm values were used as continuous predictors in later statistical analyses. Descriptive statistics for the idiom familiarity norm can be viewed in Table 10.

We selected 64 idioms that had comprehensible literal interpretations, that scored between 1 and 3 in the familiarity norm, and that had dominant (>50%) figurative meanings to use in the experiment. These idioms were highly familiar to participants, and were interpretable as both literal and figurative phrases, although the figurative meanings were dominant over the literal meanings. We selected highly familiar idioms because we wanted to be sure that participants would be able to interpret them figuratively: participants might have no figurative representations for less-familiar idioms, and therefore including them in the study would not allow us to draw conclusions about figurative processing.

Table 10. Descriptive statistics for Experiment 1B norming

Norm	Mean	SD	Range
Meaning Relatedness	3.40	.91	1.75 – 5.86
Figurative Dominance	74.89	9.85	50.17 - 89.64
Idiom Familiarity	1.83	.52	1.08 - 2.92

Phrase-Word and Meaning-Word Relatedness

28 participants completed norms of phrase-word and meaning-word relatedness. These norms were intended to estimate the degree of relatedness between the actual stimuli that participants would see in the experiment. We did this in two ways. First, in the phrase-word relatedness norm, participants were shown a phrase and one of its target words and asked to rate how related they were on a scale of 1 to 7 (1 = *very related*, 7 = *very unrelated*). For example, participants might be asked “How related are the phrase ‘hit the hay’ and the word ‘pitchfork’?”. This allowed us to roughly ascertain that participants were able to see the relation between idiomatic phrases and both literal and figurative targets. We collected relatedness comparison ratings between the idiom prime phrases and their literal- and figurative-related target words, and relatedness comparison ratings between the literal control phrases and the literal- and figurative-related target words.

Second, in the meaning-word relatedness norm, participants were shown a paraphrased meaning of an idiom and one of the target words and again asked to rate how related they were one a scale of 1 to 7. We collected relatedness comparison ratings between the figurative paraphrased idiom meaning and the figurative target, and the literal paraphrased idiom meaning and the literal target. For example, participants might be asked “How related are the phrase ‘off the ground and in the air’ and the word ‘breeze’?”. Conducting the meaning-word relatedness norm was important because it allowed us to ensure that the target words were sufficiently related to their associated idiom meanings. In the phrase-word norm, there was a chance that participants would view idioms and targets as being unrelated because each idiom also had a well-formed literal interpretation: if participants interpreted the idioms literally rather than figuratively, they would judge the figurative targets to be highly unrelated to the idiomatic meaning. We therefore conducted the meaning-word relatedness norm to check that figurative target words were truly appropriate for the figurative meanings of the idioms. Descriptive statistics for each comparison can be viewed in Table 10.

Table 11. Descriptive statistics for phrase-word and meaning-word norm comparisons

Norm	Comparison	Mean	SD	Range
Phrase-word	Idiom/ Figurative Target	2.25	.79	1 - 5
	Idiom/ Literal Target	3.82	1.35	1 – 6.83
	Control/ Figurative Target	5.84	.91	2.8 - 7
	Control/ Literal Target	5.9	.79	3.8 - 7
Meaning-word	Fig. Paraphrase/ Fig. Target	2.48	.98	1.24 - 5.27
	Lit. Paraphrase/ Lit. Target	3.34	1.17	1.47 - 6

2.2.1.3 Procedure

Because Experiment 1B was completed concurrently with Experiment 1A, the same procedure and participants were used. Participants saw 208 prime-target pairs of the types described above. However, Experiment 1B differed from Experiment 1A in the length of time for which the prime phrase was displayed. Primes were displayed for 250 ms per content word. This was to ensure that participants had enough time to process the idiom, as well as to make the designs of Experiments 1A and 1B as parallel as possible. Most primes contained two content words (ex. *kick* the *bucket*, *bent* out of *shape*), but some contained only one content word (ex. at the *beginning*, around the *mountain*) and some contained three content words (ex. *need bright* *light, run* across the *new bridge*). Primes were therefore generally displayed for 500 ms, but some were displayed for 250 or 750 ms.

Results

Data were analyzed using linear mixed effects models (reaction time) and generalized linear mixed effects models (accuracy; Baayen, 2008) in the R statistical computing package (R Development Core Team, 2013; ver. 3.0.1) and using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015; ver. 1.1-7). P-values were obtained using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2016; ver. 2.0-20). Models were fit using the fullest random effects structure that would support convergence (Barr et al., 2013). Models contained fixed effects of meaning dominance, meaning relatedness, nonword filler type (high vs. low mean bigram frequency), target type, trial number, previous trial reaction time, previous trial accuracy, target word bigram frequency, target word length, number of syllables in the target word, target word concreteness, phrase-word relatedness, and idiom familiarity. Models additionally contained random intercepts of participant and item. Finally, we included random slopes of trial number, meaning dominance, meaning relatedness, target type, nonword filler type, idiom familiarity, and phrase-word relatedness within participants, and random slopes of meaning dominance and meaning relatedness within items. In cases of non-convergence, the random slopes explaining the least variance were removed until convergence was achieved. Outcome variables were reaction time to make a lexical decision and accuracy of lexical decision. All fixed effects except for trial number and previous trial accuracy were z-score transformed to aid convergence. Additionally, we used the inverse of the reaction time outcome variable and the fixed effect of previous trial reaction time, again to aid convergence.

In a preliminary analysis, we included the fixed effect of prime display time length (250 ms vs. 500 ms vs. 750 ms). However, including this variable made model convergence prohibitively difficult because there were very few trials with a prime display length time of 250 ms. The results reported below include only trials with a 500 ms prime display time.

2.2.1.4 Reaction Time

Idiomatic vs. Literal Priming Effects

The first analysis examined the effects of idiomaticity on reaction time. We removed filler trials and incorrect trials. Additionally, we removed trials that had reaction times greater than 2.5 standard deviations outside the means by participant and nonword bigram frequency list; this removed 18 trials from analysis. 1,251 total trials were analyzed. We used the fixed and random effects described above, with two adjustments. First, we did not include the fixed effects of meaning dominance and meaning relatedness because control trials did not have associated dominance and relatedness values. Second, we included the fixed effect of prime type (critical vs. control). We also included the random slopes of prime type, nonword bigram frequency, and target type within participants, and the random slope of prime type within items; more complete models did not converge. Descriptive statistics for reaction time by prime type and target type can be viewed in Table 11.

We created three comparisons of interest using contrast coding. The first comparison compared RTs in response to a subordinate meaning-related target to those in response to a dominant meaning-related target. The second comparison compared RTs when the previous trial's accuracy was correct to when the previous trial's accuracy was incorrect, using incorrect trials as the baseline. The third comparison compared RTs in the higher nonword frequency lists to the lower nonword frequency lists, using higher frequency nonwords as the baseline. The second and third comparisons served the same purposes as in Experiment 1A.

Table 12. Descriptive statistics for Ex1B idiomatic and control RTs (ms)

Prime Type	Target Type	Mean	SD
Idiomatic	Figurative	618	201.15
	Literal	611	194.68
Control	Figurative	633	225.86
	Literal	661	263.93

Model results can be viewed in Appendix A. As predicted, there was a significant effect of prime type such that participants were faster to respond after idiomatic primes than after control primes ($\hat{\beta}$=-.13; SE=.04; t=-3.42; p<.05). This represents a classic priming effect, as control primes were unrelated to the target words. There was also a significant effect of target type such that participants were slower to respond to figurative targets than to literal targets ($\hat{\beta}$=.14; SE=.07; t=2.07; p<.05). Finally, there was a significant interaction between prime type and target type ($\hat{\beta}$=.16; SE=.07; t=2.41; p<.05).

To explore this significant interaction, we looked for the effect of target type at each level of prime type. Model estimates for all fixed effects and interactions within each prime type can be viewed in Appendix A. For the model looking at the "idiom" level of prime type, we included the random slopes of nonword bigram frequency and target type within participants; more complete models did not converge. Although there were several significant effects of control variables, there were no effects of study variables, and no interactions.

For the model looking that the "control" level of prime type, there was a marginal effect of target type such that participants were slower to respond to figurative targets than to literal targets ($\hat{\beta}$=.27; SE=.14; t=1.89; p=.06). This effect is unexpected, but theoretically uninteresting: there is no reason we would expect to see faster RTs for either target type after unambiguous control phrases, and we know from the phrase-word meaning norm that control phrases were not related to either literal or figurative targets. There were no other effects, and no interactions.

Meaning Dominance and Meaning Relatedness

The second analysis investigated effects of meaning dominance and meaning relatedness on RT. Control trials involved unambiguous words that did not have associated dominance and relatedness values, and were therefore removed from analysis. 1,205 total trials were analyzed. Descriptive statistics for reaction time by meaning dominance, meaning relatedness, and target type can be viewed in Table 12. Although meaning dominance and meaning relatedness were treated categorically in analyses, for ease of presentation they appear in Table 13 using a median split.

Table 13. Descriptive statistics for Ex1B reaction times (ms)

Target Type	Meaning Dominance	Meaning Relatedness	Mean	SD
Figurative	High	High	601	171.80
		Low	636	232.67
	Low	High	592	183.76
		Low	658	203.93
Literal	High	High	622	173.43
		Low	635	226.94

Low	High	584	169.29
	Low	607	201.13

The model contained the random slopes of target type and familiarity within participants, and meaning dominance within items; more complete models did not converge. Model estimates for all fixed effects and interactions can be viewed in Appendix A.

Although there were several significant effects of control variables, there were no significant effects of study variables, and no interactions.

2.2.1.5 Accuracy

Accuracy results grouped by Prime Type (ambiguous/control) and Target Type (literal/figurative) can be viewed in Table 13.

Table 14. Experiment 1B accuracy results

Prime Type	Target Type	Proportion Correct
Idiom	Figurative	.97
	Literal	.96
Control	Figurative	.93
	Literal	.95

Idiomatic vs. Literal Priming Effects

The first analysis examined the effects of ambiguity on accuracy. Data were trimmed as described in the reaction time analyses above. Because participants made very few errors, we transformed the data with the empirical logit, grouped over participants, to better allow models to converge. The models contained fixed effects of prime type (critical vs. control), target type (literal vs. figurative), nonword bigram frequency (high vs. low) target word length, target word number of syllables, and target word concreteness. Correct responses were coded as "hits" and incorrect responses were coded as "misses".

There was a significant effect of nonword bigram frequency such that participants made fewer errors when filler nonwords were lower frequency than they did when filler nonwords were higher frequency ($\hat{\beta}=.05$; $SE=.01$; $t=3.95$; $p<.05$). There was also a significant interaction between nonword bigram frequency and target type ($\hat{\beta}=.05$; $SE=.02$; $t=2.04$; $p<.05$). The effect of nonword bigram frequency was greater when targets were figurative. However, this interaction is difficult to interpret given that it collapses over idiomatic and control primes. This interaction can be viewed in Figure 2. Model results can be viewed in Appendix A.

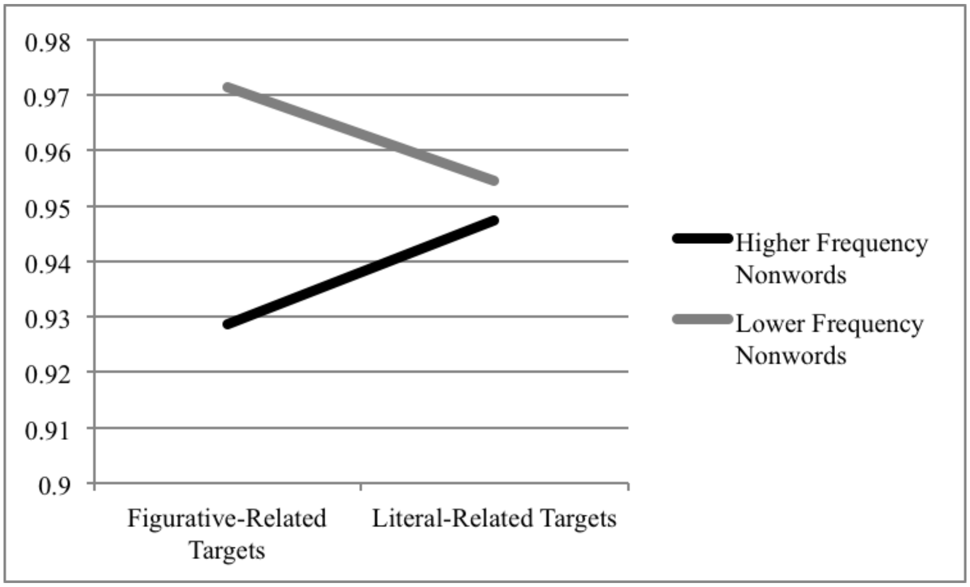

Figure 2. Effects of Target Type and Nonword Bigram Frequency on Accuracy in Ex 1B Priming

To further explore the interaction between Target Type and Nonword Bigram Frequency, we looked for the effect of Target Type within each level of Nonword Bigram Frequency. Model results for all fixed effects and interaction within each level can be viewed in Appendix A.

For higher nonword bigram frequencies, we constructed an empirical logit model using the same parameters and fixed effects described above. There was a significant effect of prime type such that participants were more accurate when the prime was an idiom compared to when the prime was a control phrase ($\hat{\beta}$=.32; SE=.09; t=3.27; $p<.05$). This suggests that the higher semantic engagement elicited by higher bigram frequency nonwords facilitates idiom processing. There were no other significant effects, and no interactions.

For lower nonword bigram frequencies, there were no significant main effects and no interactions.

Meaning Dominance and Meaning Relatedness

The second analysis investigated effects of meaning dominance and meaning relatedness on accuracy. Control trials involved unambiguous literal phrases that did not have associated dominance and relatedness values, and were therefore removed from analysis. Data were trimmed as described in the reaction time analyses above. Because participants made very few errors, we transformed the data with the empirical logit, grouped over subjects, to better allow models to converge. The models contained fixed effects of meaning dominance, meaning relatedness, target type (literal vs. figurative), nonword bigram frequency (high vs. low), target word length, target word number of syllables, and target word concreteness. Correct responses were coded as "hits" and incorrect responses were coded as "misses".

There was a significant interaction between target type and meaning dominance ($\hat{\beta}=.31$; $SE=.14$; $t=2.24$; $p<.05$). This interaction can be viewed in Figure 3. The effect of meaning dominance was greater for literal targets than for figurative targets. This is unsurprising given that in this experiment, increased dominance means increased figurativity. This suggests that, for idioms with dominant figurative meanings, the figurative meaning was accessed quickly, resulting in errors when a literal target was presented. There were no other significant main effects and no interactions. Model results can be viewed in in Appendix A.

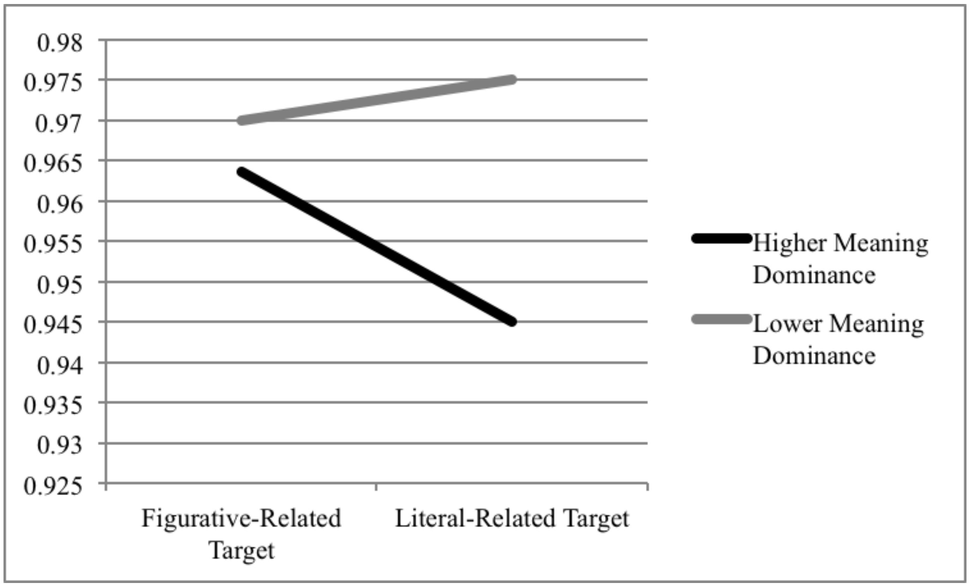

Figure 3. Effects of Meaning Dominance and Target Type on Accuracy

Discussion

The primary finding of Experiment 1B was that when the target was literal, participants were less accurate when the figurative meaning of the idiom was strongly dominant over the literal meaning (for example, when seeing the target "railing" after the idiom "on the fence"). This suggests that strongly dominant figurative meanings of idioms were accessed upon seeing the idiom prime, without any supportive or biasing context. Subsequent presentation of a literal target was incongruent with the activated idiom meaning, resulting in lowered accuracy.

Immediate initial activation of figurative meanings is characteristic of some older models of idiom processing, such as Gibbs's Direct Access Model (1980), in which stored figurative meanings are activated immediately upon encountering an idiom; compositional literal analysis happens only later, if the figurative meaning is inappropriate. However, the inverse effect of figurative dominance on accuracy observed in the present experiment is also congruent with models such as Titone and Connine's Hybrid Model (1999), in which compositional literal analysis and figurative meaning retrieval occur simultaneously. Under their model, the conventionality of an idiom—the degree to which a particular string of words is likely to have an idiomatic meaning in a particular environment, and a reasonable proxy for the dominance of the figurative meaning—influences idiom comprehension such that more conventional idioms are processed faster due to easier direct retrieval of their figurative meanings. If idioms are processed as the Hybrid Model predicts, we would expect that the idiomatic meaning of more figuratively-dominant idioms would be activated quickly upon encountering the idiom even without context, resulting in errors when a literal target is presented, as observed in the current experiment.

Finally, participants responded more quickly after seeing idiom primes than they did after seeing unambiguous control primes, although as in Experiment 1A there was no interaction with target type. For the accuracy outcome variable, participants again were more accurate following idiom primes than control primes, but only when filler nonwords were higher frequency. This provides some evidence that the higher semantic engagement elicited by the higher-frequency nonwords facilitated idiom processing in some way. This result is particularly interesting when considered in light of Cacciari and Tabossi's Configuration Hypothesis (1988). Under the Configuration Hypothesis, idiom recognition occurs when the processor recognizes a particular "configuration" of words as belonging to an idiom. Critically, semantic processing is not required for idiom recognition under this model; instead, idiom recognition may proceed entirely based on word co-occurrences, although Cacciari and colleagues do not specify either way. In the present study, we show that idiom processing is facilitated under conditions of higher semantic engagement. This suggests that, although word co-occurrences may be enough to trigger idiom recognition on their own, this process is facilitated when semantics are involved.

In sum, the results of Experiment 1B are consistent with hybrid models of idiom processing, in which retrieval of stored figurative meanings and compositional literal analysis occur simultaneously.

2.3 DISCUSSION

Taken together, the results of Experiment 1A and 1B are not similar enough to show clear parallels between ambiguous word and idiom processing, although neither do they show that these two types of ambiguity are processed differently. Rather, each experiment yielded result patterns that, although not wholly as expected, are broadly consistent with our predictions generated from previous research.

Given the extensive literature on ambiguous word processing, it is worth considering why we did not find all the effects we predicted in Experiment 1A. As touched on in the Experiment 1A Discussion, it seems likely that the lack of predicted effects was due to experimental design choices. Many studies of ambiguous word processing include unrelated control targets in addition to dominant- and subordinate-related targets (Eddington & Tokowicz, 2015). This manipulation allows researchers to look for processing facilitation and inhibition of target words relative to control targets. Advantages for polysemous words and disadvantages for homonymous words are usually found in comparison to unambiguous control words, rather than by comparing the two types of ambiguous words. Not including unrelated control targets in the present study may have made finding these effects more difficult, especially when looking for interactions with meaning dominance. It's also possible that using the ambiguous units as primes instead of targets may have obscured our ability to find the effects we sought, given that we did not actually measure processing on the ambiguous unit itself; we return to the point in the General Discussion.

Another factor that may have obscured the results of Experiment 1A is the semantic engagement elicited by the task. Armstrong and Plaut (2016) proposed that the appearance of a polysemy advantage or a homonymy disadvantage is related to semantic engagement: when semantic engagement is high, homonyms are disadvantaged compared to unambiguous words, and when semantic engagement is low, polysemes are advantaged compared to unambiguous words. Manipulating the average frequency of the nonword fillers has a similar effect on processing as changing the ISI, and results in responses being recorded at different points during the time-course of meaning activation. Our higher and lower frequency filler conditions had significantly different average frequencies, indicating that higher-frequency nonwords were more "wordlike" than lower-frequency nonwords. In spite of this, it's possible that the differences between homonyms and polysemes weren't great enough to appear statistically at the points at which we measured processing. Additionally, manipulating semantic engagement by changing nonword frequency is notoriously difficult (B. Armstrong, personal communication, July 15, 2016); it's possible that the unexpected facilitation for polysemy under conditions of high semantic engagement in the present study is simply due to small design differences between the present study and that conducted by Armstrong and Plaut.

It is likely that experimental design choices also underlie the lack of effects observed for Experiment 1B. The particular timing of the priming paradigm may have obscured any effects of meaning relatedness we might otherwise have found. In several experiments, Caillies and colleagues (2007; 2011) presented participants with either idioms or literal phrases embedded in sentences and then asked them to perform a lexical decision task on words related to the idioms' figurative meanings. Critically, they manipulated the interstimulus intervals before the target words were presented to investigate the time course along which figurative meanings became active. They consistently found that idiomatic meanings were active sooner than literal meanings, as indexed by faster reaction times following idiomatic primes at shorter ISIs. This effect varied based on the decomposability of the idiom: figurative meanings of decomposable idioms were available fastest, followed by figurative meanings of nondecomposable idioms. Critically, however, they found that this effect either weakened (Caillies & Butcher, 2011) or was absent (Caillies & Declercq, 2011) at longer ISIs, such that decomposable and nondecomposable idioms had similar reaction times. In the present study, idioms were displayed for 500 ms followed by a 250 ms ISI before the target word was presented giving participants a total of 750 ms to process the idiom. It's possible that this was long enough for participants to process the idioms completely enough to not show effects of meaning relatedness even where they would be predicted. Although we made these timing choices to make Experiment 1A and 1B comparable, it is possible that in doing so we eliminated the possibility of observing the effects we predicted for Experiment 1B.

Experiment 1 was intended to investigate potential parallels between idiom and ambiguous word processing. To do this, we conducted parallel lexical decision experiments, using ambiguous units as primes and looking for processing facilitation on target words related to the different meanings of each ambiguous unit. Although we did not find notably similar patterns of processing between idioms and ambiguous words, we also did not find notably different patterns of processing. The results of Experiment 1 therefore do not allow us to make a firm conclusion regarding whether literal and figurative language are treated the same during language comprehension. Rather, the question remains open.

Experiment 2 investigates this question from a different angle, using eyetracking during reading to examine processing of idioms embedded in supportive context sentences.

3.0 EXPERIMENT 2

In Experiment 1, we did not find definitive evidence that the language comprehension system treats ambiguous units similarly, regardless of their grain size or literality. However, neither did our results provide conclusive evidence that these two types of ambiguity are processed differently. In Experiment 2, we investigate the same questions as in Experiment 1 using a different method: eyetracking during reading. To do this, we situate idioms in sentences following either literally- or figuratively-biasing context sentences. This method has several advantages over the priming paradigm used in Experiment 1. First, it allows us to observe idiom comprehension in a more natural task and compare the results to previous studies of ambiguous word reading. Second, it allows us to more carefully examine the time course of idiom comprehension using fine-grained eyetracking data: eyetracking measures reflect the full time course of meaning activation rather than the arbitrary time slices examined in Experiment 1. Rather than running two parallel studies of ambiguous words and idioms, as we did in Experiment 1, in Experiment 2 we draw on reading studies of ambiguous words in context and conduct a similar experiment with idioms.

In Experiment 2, we examine eye movements during reading of idioms embedded in meaning-biasing context sentences, paralleling the designs used in Foraker and Murphy's (2012) and Brocher, Foraker, and Koenig's (2016) studies of ambiguous word processing. As in the present study, both Foraker and Murphy (2012) and Brocher and colleagues (2016) presented ambiguous units after sentence contexts that biased one meaning over another. Brocher and colleagues (2016) embedded ambiguous words in biasing contexts and examined how eye movements changed when the ambiguous words were homonyms versus polysemes. They found overall more disruption to eye movements in response to homonyms than to polysemes, although they did find some early disruption for polysemes when the polyseme was preceded by disambiguating context. These results parallel the polysemy advantage found in priming studies, and suggest that increased meaning relatedness aids comprehension in both meaning activation and meaning integration. Foraker and Murphy (2012) also embedded ambiguous words in biasing contexts and found clear advantages for processing of dominant meanings of both homonyms and polysemes, even when the ambiguous word was preceded by a neutral context condition that biased neither meaning. They interpreted these results as suggesting that dominant meanings of ambiguous words are immediately available and accessed automatically. Critically, Foraker and Murphy (2012) also found an interaction between dominance scores and meaning relatedness such that polysemes with one highly dominant sense and otherwise highly semantically related senses were harder to process than more balanced polysemes. This indicates that high meaning dominance can inhibit the processing advantage usually seen for words with highly related meanings, causing them to be processed more like homonyms, which have low semantic relatedness between meanings.

Following Brocher and colleagues (2016) and Foraker and Murphy (2012), in Experiment 2 we present participants with idioms embedded into context biasing either their literal or figurative interpretations. Each item consists of a biasing context sentence, a sentence containing an idiom, and a wrap-up sentence. However, our items differ from Foraker and Murphy's and Brocher and colleagues' in one critical way: following the biasing context, our items are otherwise neutral, and the idiom sentence and the wrap-up sentence are congruent with both interpretations of the idiom. Therefore, the context sentence is the only determiner of which meaning of the idiom is activated. This enables us to examine if and how biasing context interacts with the dominant figurative meaning of an idiom to drive interpretation. It also allows us to observe whether changes in context bias result in comprehension facilitation or difficulty in neutral regions following the idiom. In contrast, Foraker and Murphy included a disambiguating region that established exactly which meaning—dominant or subordinate—of the ambiguous words was intended. Brocher and colleagues used a context location manipulation to disambiguate their ambiguous words: biasing context appeared either before or after the critical word. Although these contextual manipulations mean that we may see slightly different patterns of results, both Foraker and Murphy and Brocher and colleagues found effects on their critical words, and we therefore expect to see effects on the idioms in our study even without disambiguating regions or manipulations of context location.

Under a view in which idioms and ambiguous words are treated similarly by the language processing system, we predict there will be greater eye movement disruption to idioms with less related meanings compared to idioms with more related meanings, indicating increased processing difficulty as the literal and figurative meanings of idioms become less related. This prediction is congruent with both research on idiom decomposability (Caillies & Butcher, 2007; Titone & Connine, 1999) and research showing an advantage for polysemous words in both priming and reading studies (Klepousniotou, 2002). One explanation for the easier processing of decomposable idioms is that their literal and figurative meanings are closely related, similar to polysemes. We also expect to see more disruption to comprehension when the context biases the literal, subordinate, meaning of the idiom, in line with Foraker and Murphy's (2012) finding of increased difficulty processing ambiguous words following a subordinate-biased context. However, we expect that this effect will interact with meaning relatedness: idioms with lower semantically related meanings will show more disruption after literally-biasing sentences than will idioms with highly semantically related meanings, following Brocher and colleagues' (2016) finding that processing difficulty following subordinate-biased context was greater for homonyms than for polysemes. Finally, we predict that idioms with high figurative dominance and overall highly semantically related meanings will show more disruption to eye movements compared to idioms with more balanced meanings, paralleling patterns seen in ambiguous word comprehension (Foraker & Murphy, 2012). Finding these interactions will be suggestive evidence that ambiguities are resolved similarly regardless of their literality.

3.1 METHODS

Materials

Participants read three-sentence passages while their eyes were tracked. All items can be viewed in Appendix B. Passages consisted of a biasing context sentence, a sentence containing an idiom, and a short wrap-up sentence. Context sentences biased either the literal or figurative interpretation of the idiom. There were 45 items with two conditions each, as shown below:

(1A) Peter often panics about small things. *Sometimes he* <u>loses his grip</u> [for a moment.] But he always recovers. (figuratively-biased context)

(1B) Peter is an expert on the parallel bars. *Sometimes he* <u>loses his grip</u> [for a moment.] But he always recovers. (literally-biased context)

The only portion of each passage that varied was the context sentence. Idiom sentences consisted of a neutral precritical region (italicized above), an idiom (underlined above), and a neutral postcritical region (in brackets above). Both precritical and postcritical regions were at least two words long to increase the likelihood that they would be fixated. Idiom and wrap-up sentences were consistent with both literal and figurative interpretations of the idiom, and therefore required the biasing context sentence to guide interpretation. Stimulus presentation was counterbalanced such that each participant only saw one of the two possible contexts for each item. Each sentence was presented on its own line, and the critical idiom region was always in the middle of the second line of text.

Additionally, 80 filler items were included. These items were three-sentence passages of varying sentence lengths and syntactic structures (for example, “Aelita is wild about marine biology. She studies sea cucumbers and tells her friends all about them. Who knew sea cucumbers were so interesting?”). Some filler passages contained figurative language such as metaphor, simile, or hyperbole, and these were distributed across the three sentences. After 40 of the filler passages a yes/no comprehension question appeared, half of which required a “yes” response. Where possible, comprehension questions related to information provided in the third sentence of the passage to increase the likelihood that it would be fixated.

Norming

Context bias

To confirm that the context sentence biased the appropriate meaning of the idiom, we conducted a context bias norm. Twenty-four undergraduate students from the University of Pittsburgh participated. Participants were presented with each passage truncated after the idiom and asked to rate on a scale of 1 to 7 which meaning of the idiom, literal or figurative, was likely intended. A paired-sample *t*-test indicated that participants preferred the figurative meaning after reading the figuratively-biased context ($M = 1.55$; $SD = .42$) and the literal meaning after reading the literally-biased context ($M = 5.85$; $SD = .85$; $t(44) = -30.77$; $p<.05$). We reverse-coded the literally-biased context scores. This meant that a higher context bias norm score meant a more biased context, regardless of literality, and a lower score meant a less biased context. These unidirectional norm scores were used as predictors in later analyses.

Progressive naturalness

Six participants evaluated naturalness at several points in the passage to ensure that any differences in eye movements between conditions would be due to the experimental manipulations, rather than be a reaction to sudden decreases in naturalness. Participants rated the naturalness of each passage at four points: immediately before the idiom, immediately after the idiom, after the idiom sentence, and after the full item. Naturalness was rated on a scale of 1 to 7, with 1 being very natural and 7 being very unnatural. Naturalness evaluation points are indicated using (/) in (2) below.

> (2) Peter often panics about small things. Sometimes he/ loses his grip/ for a moment./ But he always recovers./

Means and standard deviations of naturalness ratings at each point are shown in Table 14. A repeated measures 4 (naturalness point) x 2 (context bias) ANOVA showed no effect of context bias ($F(1,44) = 1.32$; $p = .26$) and a significant effect of point ($F(3,132) = 66.82$; $p<.05$). There was also a significant interaction ($F(3,132) = 2.07$; $p<.05$). For both literal and figurative contexts, naturalness was at its lowest immediately before the idiom and increased immediately after the idiom. However, for figurative contexts naturalness did not change after the idiom and through the end of the idiom sentence, and decreased slightly at the end of the passage. In contrast, for literal sentences naturalness continued increasing through the end of the idiom sentence and did not change through the end of the passage. Naturalness ratings will be used as predictors in later analyses.

Table 15. Descriptive statistics for Ex 2 progressive naturalness norm scores

Context bias	Evaluation point	Mean	SD
Figurative	Before Idiom	3.74	.64
	After Idiom	2.65	.82
	After Idiom Sentence	2.48	.57
	Full Passage	2.8	.66
Literal	Before Idiom	3.76	.78
	After Idiom	2.72	.88
	After Idiom Sentence	2.29	.61
	Full Passage	2.47	.68

Procedure

Thirty-six undergraduate students from the University of Pittsburgh who had not participated in the norming completed the experiment for course credit. Before participating in the study, all participants provided informed consent and completed a questionnaire collecting demographic information such as age and language background. All participants were native speakers of English.

Following consenting and demographic collection, participants completed the eyetracking experiment. An Eyelink 1000 eyetracker monitored the gaze location of participants' right eyes every millisecond. Participants viewed stimuli binocularly on a monitor 63 cm from their eyes. Approximately three characters equaled 1º of visual angle. The experiment lasted approximately 30 minutes. Chin and forehead rests minimized head movements. Participants were asked to read normally and were told that after some sentences they would be asked a yes-no comprehension question; participants responded to comprehension questions using a mouse. The tracker was calibrated using a 9-point calibration grid before the experiment, and additional re-calibration was performed during the experiment as necessary. A single-point drift check was performed every five trials.

3.2 RESULTS

Eye movement analysis focused on the following eye movement measures: (1) *first fixation*, the duration of the first fixation on a region during first-pass reading, (2) *go past*, the sum of all fixations from entering a region during first-pass reading until leaving it to the right, including any regressive fixations, (3) *re-reading*, the sum of all fixations on a region not during first-pass reading, and (4) *total time*, the sum of all fixations on a region. In general, we focused on later eye-tracking measures in our analyses because they are often where meaning-related effects appear. However, we also looked at first fixation because it could be informative about spillover effects in the postcritical target region. We did not include first pass in our set of eye movement measures because participants would occasionally briefly fixate the second and third lines of each item when making return sweeps to the beginning of the line. Although this means that sometimes the "first fixation" on a region was actually a brief error fixation, it's possible that these short fixations did affect processing of the region, so we included them in analyses.

Trials with track losses and blinks during first pass reading of the idiom were removed (3.4% of trials). Short fixations (<60 ms) within one character position of a preceding or following longer fixation were combined. Other fixations less than 60 ms in duration were removed, as were fixations longer than 1500 ms (Brocher et al., 2016; Frisson, Harvey, & Staub, 2017; Weiss, Kretzschmar, Schlesewsky, Bornkessel-Schlesewsky, & Staub, 2017). Additionally, three trials in the literal context bias condition were removed due to calibration errors. After processing, 1,561 trials were included in analyses.

Data were analyzed using linear mixed effects models in the R statistical computing package (R Development Core Team, 2013; ver. 3.0.1) and using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015; ver. 1.1-7). P-values were obtained using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2016; ver. 2.0-20). Models included fixed effects of context bias (literal vs. figurative), meaning dominance, meaning relatedness, idiom familiarity, trial number, and unidirectional context bias norm score. For each region of interest, we also included the corresponding progressive naturalness score as a fixed effect. We included random effects of participant and item. Finally, we included random slopes of progressive naturalness score and unidirectional context bias norm score within items and meaning dominance, meaning relatedness, idiom familiarity, progressive naturalness score, trial number, and unidirectional context bias norm score within participants. Models were fit using the fullest random effects structure that would support convergence (Barr et al., 2013). To increase interpretability of the results, we reverse-coded meaning relatedness, familiarity, and progressive naturalness.

We included fixed effects of trial number and unidimensional context bias norm score in a first set of analyses. However, including these variables made model convergence prohibitively difficult, and models without these variables gave the same results and allowed us to include more random effects. We therefore report model results without inclusion of trial number and unidirectional context bias norm score fixed effects.

Means and standard deviations for all eye movement measures in all regions can be seen in Table 15.

Table 16. Descriptive statistics for all eye movement measures in all analysis regions

Region	Measure	Context Bias	Mean (ms)	SD
Precritical	First Fixation	Literal	213	76.50
		Figurative	210	81.39
	Go Past	Literal	482	333.11
		Figurative	485	354.83
	Rereading	Literal	314	201.41
		Figurative	312	231.85
	Total Time	Literal	544	370.76
		Figurative	542	401.09
Critical	First Fixation	Literal	214	65.58
		Figurative	211	69.87
	Go Past	Literal	573	301.09
		Figurative	579	293.37
	Rereading	Literal	330	192.99
		Figurative	341	210.56
	Total Time	Literal	649	340.85
		Figurative	647	344.47
Postcritical	First Fixation	Literal	226	78.33

	Figurative	229	87.99
Go Past	Literal	789	617.29
	Figurative	780	560.97
Rereading	Literal	449	412.09
	Figurative	411	340.58
Total Time	Literal	715	453.29
	Figurative	714	429.13

Pre-Critical Region

Model results for all eye tracking measures in the pre-critical region can be seen in Appendix A.

3.2.1.1 First Fixation

There were no significant effects, and no interactions.

3.2.1.2 Go Past

There were no significant effects, and no interactions.

3.2.1.3 Re-Reading

There were no significant effects, and no interactions.

3.2.1.4 Total Time

There were no significant effects, and no interactions.

Critical Region

Model results for all eye tracking measures in the critical region can be seen in Appendix A.

3.2.1.5 First Fixation

To aid convergence, we removed the random slopes of meaning relatedness and figurative dominance within participants and progressive naturalness score within items. There were no significant main effects, and no interactions.

3.2.1.6 Go Past

There was a significant effect of figurative dominance such that as figurative dominance increased, go past time decreased ($\hat{\beta}$=-254.06; SE=87.97; t=-2.89; p<.05). There was also a significant effect of meaning relatedness such that as relatedness increased, go past time decreased ($\hat{\beta}$=-547.01.42; SE=170.21.; t=-3.21; p<.05). There was a significant interaction between meaning dominance and meaning relatedness ($\hat{\beta}$=76.96; SE=23.05; t=3.28; p<.05): the effect of figurative dominance was greater when meaning relatedness was high (see Figure 4). Having a strongly dominant figurative meaning appears to inhibit processing when meaning relatedness is also high, regardless of context bias. Finally, there was a marginal effect of familiarity such that as familiarity increased, go past time decreased ($\hat{\beta}$=-71.54; SE=39.35; t=-1.82; p=.08). There were no other significant effects or interactions.

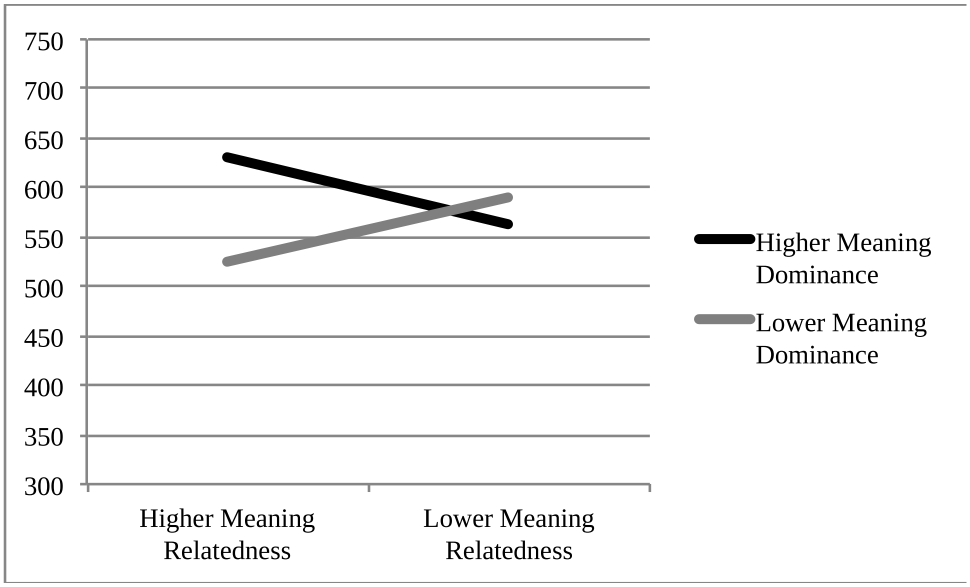

Figure 4. Effects of Meaning Dominance and Meaning Relatedness on Critical Region Go Past Time

3.2.1.7 Re-Reading

All effects that were present in Go Past also appeared in Re-Reading. There was a marginal effect of figurative dominance such that as figurative dominance increased, second pass time decreased ($\hat{\beta}$=-91.86; SE=45.61; t=-2.01; p=.051). There was also a significant effect of meaning relatedness such that as relatedness increased, second pass time decreased ($\hat{\beta}$=-197.38; SE=87.89; t=-2.25; $p<.05$). There was a significant interaction between meaning dominance and meaning relatedness ($\hat{\beta}$=26.60; SE=12.24; t=2.17; $p<.05$): when meaning dominance was high, participants read equally fast regardless of meaning relatedness (Figure 5). When meaning dominance was low, participants read faster when meaning relatedness was high. As in Go Past, lower meaning dominance appears to facilitate processing when meaning relatedness is high. Finally, there was a marginal effect of familiarity such that as familiarity increased, second pass time decreased ($\hat{\beta}$=-38.93; SE=20.33; t=-1.92; p=.06). There were no other significant effects or interactions.

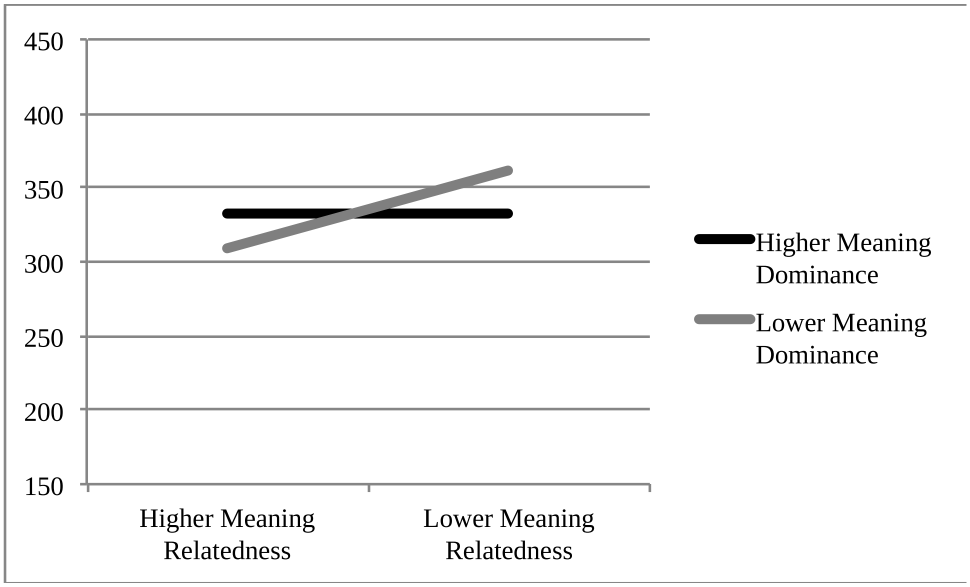

Figure 5. Effects of Meaning Dominance and Meaning Relatedness on Critical Region Re-Reading Time

3.2.1.8 Total Time

To aid convergence, we removed the random slope of progressive naturalness score within items. All effects that were present in Go Past and Re-Reading also appeared in Total Time. There was a significant effect of meaning dominance such that as figurative dominance increased, total time decreased ($\hat{\beta}$=-214.91; SE=91.01; t=-2.36; p<.05). There was also a significant effect of meaning relatedness such that as relatedness increased, total time decreased ($\hat{\beta}$=-465.92; SE=176.08; t=-2.65; p<.05). There was a significant interaction between meaning dominance and meaning relatedness ($\hat{\beta}$=65.65; SE=24.31; t=2.70; p<.05): the effect of figurative dominance was greater when meaning relatedness was high. Again, having a strongly dominant figurative meaning appears to inhibit processing when meaning relatedness is also high. This interaction can be viewed in Figure 6. Finally, there was a significant effect of familiarity such that as familiarity increased, total time decreased ($\hat{\beta}$=-91.72; SE=40.71; t=-2.25; p<.05).There were no other significant effects or interactions.

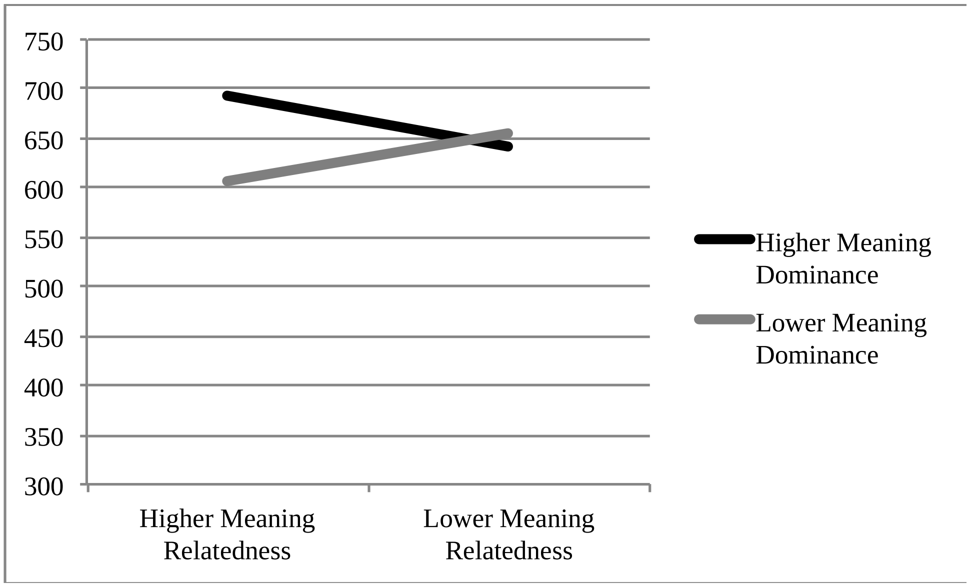

Figure 6. Effects of Meaning Dominance and Meaning Relatedness on Critical Region Total Time

Post-Critical Region

Model results for all eye tracking measures in the post-critical region can be seen in Appendix A.

3.2.1.9 First Fixation

To aid convergence, we removed the random slope of progressive naturalness within items. There was a marginal interaction between context bias and meaning dominance ($\hat{\beta}$=35.29; SE=18.81; t=1.88; p=.06). There was also a marginal interaction between context bias, figurative dominance, and meaning relatedness ($\hat{\beta}$=-9.06; SE=5.02; t=-1.81; p=.07). There were no other effects, and no interactions.

To investigate the marginal interaction between context bias, figurative dominance, and meaning relatedness, we looked for the effects of Figurative Dominance and Meaning Relatedness within each level of Context Bias. Model estimates for all fixed effects and interactions within each level of context bias can be viewed in Appendix A. For literally-biased contexts, we constructed a linear mixed effects model using the same parameters and fixed effects described above. There were no significant effects, and no interactions.

For figuratively-biased contexts, we removed the random slope of progressive naturalness score within items to aid convergence. There was a significant effect of figurative dominance such that as figurative dominance increased, first fixation time also increased ($\hat{\beta}$=34.96; SE=15.28; t=2.29; p<.05). There was also a marginal interaction between figurative dominance and meaning relatedness ($\hat{\beta}$=-7.50; SE=4.07; t=-1.09; p=.07): first fixation times were shortest when idioms were low in both meaning dominance and meaning relatedness. When context was biased towards a figurative interpretation, a strongly figuratively dominant meaning inhibited initial processing of the postcritical region. Instead, participants had shorter fixations when meaning dominance was more balanced, especially when meaning relatedness was also low. Visualization of this interaction appears in Figure 7. Note that the scale on the y-axis of this figure increases in steps of two rather than 100, as in the other figures in this section.

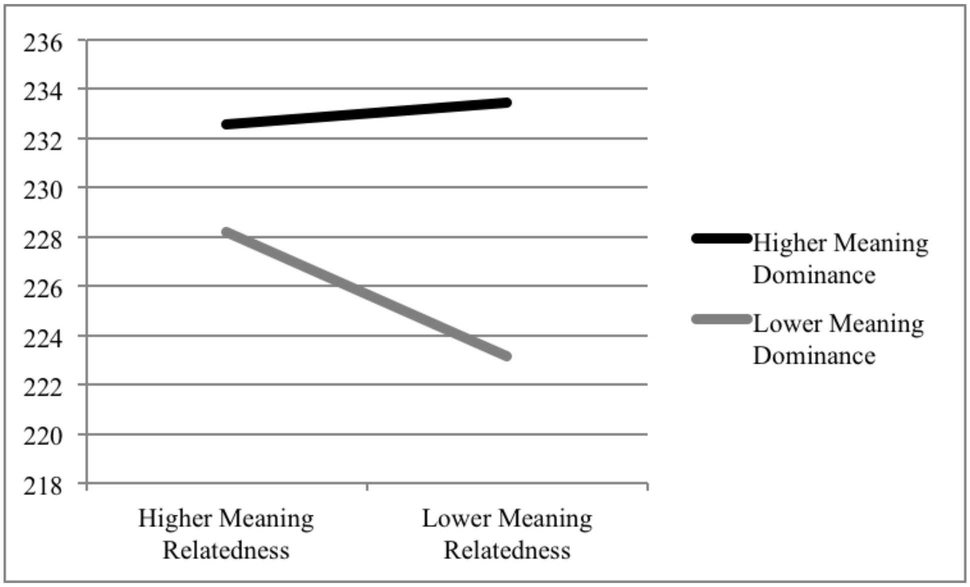

Figure 7. Effects of Meaning Dominance and Meaning Relatedness on Postcritical Region First Fixation for Figuratively-Biased Contexts

3.2.1.10 Go Past

There was a significant effect of progressive naturalness score such that as naturalness decreased, go past time increased ($\hat{\beta}$=-75.81; SE=25.00; t=-3.03; p<.05).

3.2.1.11 Re-Reading

There were no significant effects, and no interactions.

3.2.1.12 Total Time

To aid convergence, we removed the random slope of progressive naturalness score within items. There was a significant effect of progressive naturalness score such that as naturalness decreased, total time increased ($\hat{\beta}$=-39.47; SE=17.62; t=-2.24; p<.05)

3.3 DISCUSSION

In the current experiment, we found that figurative dominance and meaning relatedness differently affected participants' eye movements during idiom comprehension. The majority of dominance and relatedness effects were seen in the critical region: participants were faster to read idioms with highly dominant figurative meanings, as well as idioms with highly related literal and figurative meanings, in both earlier (go past) and later (re-reading; total time) eye movement measures. However, these effects also interacted to drive eye movements. In both go past and total time, the effect of figurative dominance was greater when meaning relatedness was high: participants read idioms slowest when the figurative meaning was strongly dominant and the literal and figurative meanings were highly related (*on the fence*), and fastest when literal and figurative meanings were highly related but the figurative meaning was not dominant over the literal *(play with fire*).

The interaction between figurative dominance and meaning relatedness is strongly suggestive evidence that ambiguous units are resolved similarly regardless of their literality or grain size. Foraker and Murphy (2012) found processing disadvantages for ambiguous words with one highly dominant sense and otherwise highly related senses (*gem*). This interaction is strikingly similar to that observed in the present study, in which highly figuratively dominant idioms with highly related literal and figurative meanings were disadvantaged in processing compared to idioms with highly related but more balanced meanings. Foraker and Murphy explained this interaction by suggesting that, when both sense similarity and sense dominance are strong, increased competition results in increased difficulty for comprehenders. This explanation also makes sense when applied to the present results: although idioms with highly related meanings are usually advantaged in processing specifically because their literal and figurative meanings are so closely related, this closeness results in interference from concurrent literal analysis when the figurative meaning is strongly dominant. The fact that the same constructs not only have similar effects on both idioms and ambiguous words but also interact to drive processing in similar ways suggests that the language system treats idioms and ambiguous words similarly during processing.

The advantage in reading times for idioms with highly related literal and figurative meanings—what might be termed "polysemous" idioms—is congruent with previous research on both idiom comprehension and ambiguous word processing. This congruence supports the idea that transparency or decomposability in idioms and meaning relatedness in ambiguous words are likely the same construct. Studies of decomposable idioms—idioms with more related literal and figurative meanings, analogous to polysemes—have found earlier activation for figurative meanings of decomposable idioms (Caillies & Butcher, 2007; Caillies & Declerq, 2011), shorter reading times for decomposable idioms (Gibbs, Nayak, & Cutting, 1989), and facilitated processing for both literal and figurative meanings of decomposable idioms (Titone & Connine, 1999). It seems that stronger relationships between literal and figurative idiom meanings confer advantages at many levels of processing. Critically, similar results have been found for polysemes; for example, in a study of reading times Brocher and colleagues found less difficulty for polysemes compared to unambiguous words, and Klepousniotou and colleagues found advantages for polysemes compared to homonyms using both behavioral and ERP measures (Klepousniotou, 2002; Klepousniotou & Baum, 2007; Klepousniotou et al., 2012; Klepousniotou et al., 2008). The consistency in the advantage for "polysemous" ambiguous units across experimental measure and ambiguous unit grain size is striking, and suggests that meaning relatedness, transparency, and decomposability all describe the same construct, which has similar effects on ambiguity resolution regardless of the literality or grain size of the ambiguous unit.

Notably, the only effect of context bias was on first fixation duration in the postcritical region, where context bias interacted with both figurative dominance and meaning relatedness; otherwise, no effects of context were observed in any region on any eyetracking measure. This is unexpected given that the contexts strongly biased interpretation of the idioms towards literal or figurative: based on ambiguous word research, we expected that context bias would interact with figurative dominance, resulting in disrupted processing when a highly figuratively dominant idiom appeared after a literally-biased context (Brocher et al., 2016). However, the absence of context effects is consistent with some previous work on idiom comprehension. Titone and Connine (1999) did not find effects of context bias, whether literal or figurative, for either decomposable or nondecomposable idioms. Instead, they observed differences in processing depending on whether the context was located before the idiom, thereby biasing it, or after the idiom, thereby disambiguating it. Interestingly, effects of context bias often appear within a context location manipulation in studies of ambiguous words. Brocher and colleagues (2016) found different processing disadvantages depending on whether a word was homonymous or polysemous and whether context biased its dominant or subordinate meaning, but these effects also interacted with the location of the biasing context. Similarly, Foraker and Murphy (2012) found effects of dominant or subordinate contexts, but these effects manifested in a congruent or incongruent disambiguation region, not on the ambiguous word itself. It's possible that context bias effects were not observed in our study because the context sentence served both to bias the idiom and to disambiguate it, and we would have needed to include an explicitly disambiguating region to see any context effects. Because our postcritical and wrap-up regions were as neutral as possible and didn't vary depending on context conditions, it's possible that comprehending them did not force participants to process the idioms, whether literally or figuratively, as deeply as they might have otherwise.

In sum, the results of the present experiment are largely congruent with previous research on both idiom and ambiguous word comprehension, and support a language comprehension system that is flexible when resolving both literal and figurative ambiguities at multiple grain sizes. Moreover, the parallels between the advantages seen for polysemes in ambiguous word processing and "polysemous" idioms in the present and other studies suggests that meaning relatedness, decomposability, and transparency are all slightly different ways of operationalizing the same underlying construct.

4.0 GENERAL DISCUSSION

Across two experiments, we investigated ambiguity resolution at multiple grain sizes by examining the effects of meaning dominance and meaning relatedness on ambiguous word and idiom comprehension. In two parallel sub-experiments, Experiment 1 compared facilitation for dominant and subordinate meanings of ambiguous units in a primed lexical decision task. For ambiguous words, we found increased facilitation for homonyms as meaning dominance increased, for both accuracy and reaction time. This effect is as we expected and is consistent with previous research on ambiguous word processing (Klepousniotou et al., 2008). However, our results for polysemes are harder to interpret: we found facilitation of accuracy for polysemes when semantic engagement was high, contrary to what we predicted. Additionally, we found reduced reaction time for polysemes when semantic engagement was high and meaning dominance was also high, contrary to our predictions for the effects of both meaning dominance and semantic engagement. However, it's possible that our experimental design choices at least partly explain these unusual effects.

For idioms, we also found effects of dominance: when the target was literal, participants were less accurate when the figurative meaning of the idiom was strongly dominant over the literal meaning. This result indicates that figurative meanings of idioms were accessed upon seeing figuratively-dominant idiom primes in isolation, without any biasing or supportive context (Gibbs, 1980).

Although the results of both sub-experiments of Experiment 1 were broadly consistent with both our predictions and previous research on idiom and ambiguous word comprehension, they do not allow us to say conclusively that the language comprehension system treats ambiguous words and idioms similarly during ambiguity resolution. However, neither do they provide evidence that these two types of language are clearly processed differently. To continue our investigation of these questions, we conducted a second experiment using eyetracking during reading. This allowed us to examine how biasing context affected idiom meaning activation, as well as how idiom meanings were integrated into a larger text.

In Experiment 2, we found facilitative effects of both figurative dominance and meaning relatedness on processing of the idiom. These two constructs also interacted to drive comprehension: participants read the idioms slowest when both figurative dominance and meaning relatedness were high, and fastest when meaning relatedness was high and figurative dominance was low. Notably, this interaction is similar to one that Foraker and Murphy (2012) observed for ambiguous words in a similar eye-tracking study, and is suggestive evidence for a language comprehension system that resolves ambiguities similarly regardless of grain size or literality. Additionally, the facilitative effect of meaning relatedness in idioms parallels the polysemy advantage observed for ambiguous words, and suggests that decomposability, transparency, and meaning relatedness may all describe the same construct, which is universal to ambiguity resolution again regardless of grain size or literality.

4.1 COMPARISON OF EXPERIMENTS 1B AND 2

Notably, not only did we find different effects for idioms and ambiguous words in Experiments 1A and 1B, we also found different patterns of idiom processing between Experiments 1B and 2. In Experiment 1B, we did not find facilitative effects of figurative dominance or meaning relatedness; instead, we found general facilitation for idioms compared to control primes, and hints that higher semantic engagement might aid idiom comprehension. In contrast, in Experiment 2 we found faster reading times with increased figurative dominance and meaning relatedness, as well as slower reading times when both figurative dominance and meaning relatedness were high.

One possible explanation for these different results is that design differences impaired our ability to find similarities across experiments. As mentioned in the Discussion for Experiment 1B, the timing of the priming paradigm may have made observation of meaning relatedness effects especially difficult to observe: previous research has found that processing differences between decomposable and nondecomposable idioms appear very early after prime offset in priming paradigms (Caillies & Butcher, 2007; Caillies & Declerq, 2011), and these differences disappear with longer ISIs. It's possible that, in the current experiment, participants were given enough time to sufficiently activate both literal and figurative meanings of the idiom (occurring simultaneously under the Hybrid Model), resulting in facilitation regardless of whether the target was literal- or figurative-related. Shortening the amount of processing time available to participants in Experiment 1B may make any effects of meaning relatedness more prominent.

Another design difference that may have produced different effects in Experiments 1B and 2 is that our outcome variables were collected from different parts of the task in each experiment. In Experiment 1B, reaction time and accuracy were collected from target words, not the idiom itself. Using ambiguous units as primes rather than targets allowed us better control of our manipulation. However, it meant that, rather than collecting data during actual idiom processing, our data instead measure downstream effects of idiom characteristics on meaning facilitation. Additionally, we did not include unrelated control target words. Although effects of meaning dominance and meaning relatedness were not visible on related targets alone, they may have appeared in comparison with unrelated control target words.

In contrast, Experiment 2 had almost the opposite design as Experiment 1B: we preceded each idiom with context meant to bias interpretation toward the literal or figurative, and measured processing of the idiom itself in response to said contextual biases. It's possible that figurative dominance and meaning relatedness may have different effects on initial processing of idioms compared to downstream processing following idioms. However, testing this possibility would require either a contextual location manipulation or inclusion of a disambiguation region, rather than the neutral post-idiom regions used in Experiment 2. The addition of a disambiguation region might also address the lack of context bias effects observed in Experiment 2: many studies of both idiom comprehension and ambiguity resolution in reading see context bias effects on later disambiguation regions rather than on the ambiguous unit itself.

4.2 FUTURE RESEARCH

There are several design choices we could make in future experiments that could allow us to better compare idiom and ambiguous word processing. A critical avenue of investigation for future priming studies is to determine the appropriate timing choices to make when looking for effects of meaning dominance and meaning relatedness in idioms. Therefore, comparing target word facilitation at multiple ISIs following idioms is a crucial next step in this line of research.

Another interesting avenue for further research is to better determine the effects of higher and lower semantic engagement on idiom processing. The present study found hints of facilitation of idiom processing under conditions of higher semantic engagement during priming. However, although we did experimentally manipulate semantic engagement by changing the average frequencies of the nonword filler targets, a full investigation of semantic engagement during idiom processing was not the goal of the present study. Instead, our semantic engagement manipulation was intended to strengthen our comparison between idioms and ambiguous words by providing another arena in which to look for processing similarities. This manipulation was also intended to account for different processing effects appearing at different ISIs (as in Caillies & Butcher, 2007; Caillies & Declerq, 2011). This allowed us to measure processing at different points in the time-course of meaning activation while keeping the ISI the same, thereby giving us better control of our manipulation. A deeper investigation of semantic engagement during idiom processing could yield valuable information about how idioms are represented and processed.

Future studies of idiom processing using eye-tracking during reading would benefit by more closely mirroring the designs used in studies of ambiguous word processing during reading. In particular, including a context location manipulation would likely show effects of idioms on comprehension that were not visible in the neutral post-idiom regions used in the present study. We did not include this manipulation in Experiment 2 because were interested in processing changes on the idiom itself following a biasing context sentence. However, putting a biasing sentence after the idiom forces the reader to commit to a particular idiom interpretation, potentially resulting in disruption when the automatically activated meaning and the meaning biased by the sentence are in conflict. This effect should also interact with an idiom's figurative dominance and the semantic relationship between the idiom's meanings, as predicted by the Hybrid Model (Titone & Connine, 1999). Manipulating the location of a biasing context would also allow investigation of the processing differences between idiom facilitation when context precedes the idiom, and idiom disambiguation when context follows the idiom.

Finally, further research should investigate other ways idiom comprehension can be integrated into language comprehension as a whole. One particular avenue worthy of further investigation is examining potential processing similarities between idioms and literal multiword phrases. Preliminary investigation into the question has shown effects of both frequency and meaning on processing of idioms and matched literal control phrases, suggesting that idioms and meaningful frequency-matched literal phrases are processed similarly once frequency is controlled for (Jolsvai, McCauley, & Christiansen, 2013). Further investigation into this topic could allow researchers to both refine theories of idiom representation and develop theories of potential literal multiword representation.

4.3 IMPLICATIONS FOR MODELS OF IDIOM COMPREHENSION

Several key findings of the present study are consistent with existing models of idiom representation and processing. As described in the Introduction, models of idiom comprehension can be roughly divided into three types: *noncompositional* models, in which idioms are stored as single wordlike units which are directly retrieved during comprehension; *compositional* models, in which compositional analysis of the idiom's words is necessary for comprehension of both literal and figurative meanings; and *hybrid* models, in which compositional literal analysis and direct figurative retrieval happen simultaneously. Although the experiments in the present study were not designed to adjudicate between individual models, our results have implications for several specific models of idiom comprehension. In particular, we will discuss our results in light of noncompositional and hybrid models.

We will first turn to an examination of noncompositional models of idiom comprehension. There are two representative examples of this type of model: Swinney and Cutler's Lexical Representation Hypothesis (1979) and Gibbs's Direct Access Model (1980; 1986). Under the Lexical Representation Hypothesis, the figurative meanings of idioms are stored in the lexicon as though they were simply long words. Presentation of the first word of an idiom immediately triggers both automatic retrieval of the figurative meaning and compositional analysis of the literal meaning. Because retrieving one long word (the figurative meaning) is easier than compositional analysis of the literal meaning, the figurative meaning is processed first and has priority in comprehension. Gibbs's Direct Access Model also gives figurative meanings priority over literal meanings. However, under this model there is no simultaneous compositional analysis of the idiom's literal meaning. Instead, literal analysis happens only if the idiomatic meaning is found to be inappropriate for the sentential context after retrieval.

Both the Lexical Representation Hypothesis and the Direct Access Model predict that idiomatic meanings will be accessed faster than literal ones, although for different reasons. In the present study, the lack of a context effect on reading times in Experiment 2 is particularly contradictory to the predictions of the Direct Access Model. Under this model, idioms in figurative contexts should be processed faster than idioms in literal contexts because the idiomatic meaning is the only one initially accessed upon presentation of an idiom: the process of comparing the idiomatic meaning to the literal context, detecting a mismatch, and performing a literal reanalysis should result in slower reading times when context is literally-biased. In contrast, equally fast reading times regardless of context could be observed under the lexical representation hypothesis because literal and figurative analysis happen simultaneously. The results of Experiment 2 are therefore more consistent with the Lexical Representation Hypothesis.

However, we also found similar effects of figurative dominance and meaning relatedness on idiom reading times regardless of context bias, suggesting that figurative processing proceeds even when it's not necessary for comprehension. It's unclear whether this pattern of results would be predicted under the Lexical Representation Hypothesis, but does support the idea that figurative meanings of idioms are being stored as lexicalized "chunks", akin to long words.

Two representative examples of hybrid models of idiom comprehension are Cacciari's Configuration Hypothesis (Cacciari & Glucksberg, 1991; Cacciari & Tabossi, 1988) and Titone and Connine's Hybrid Model (1999). Under the Configuration Hypothesis, interpretation proceeds literally until the comprehender recognizes the configuration of words that they are reading corresponds to an idiom. At this point the figurative meaning of the idiom is directly retrieved, and literal analysis does not continue. Importantly, identification of the idiom is driven by co-occurrence frequencies of the words in the idiom rather than any semantic relationship between the words in the idiom and the idiom's figurative meaning. The most important construct affecting comprehension is therefore familiarity of the phrase: the more familiar the phrase is to a comprehender, the easier it is to recognize the configuration. Supporting this model, Tabossi, Fanari, and Wolf (2009) found equally fast meaningfulness judgments for decomposable idioms, nondecomposable idioms, and compositional clichés, and concluded that familiar phrases are recognized faster than unfamiliar phrases regardless of their idiomaticity. However, Smolka and colleagues (2007) found activation for the literal meanings of German verbs even when they appeared at the ends of figuratively-biased phrases, suggesting that literal analysis continues even after the figurative phrase has been identified, contrary to the predictions of the Configuration Hypothesis.

Titone and Connine's Hybrid Model, in contrast, proposes simultaneous compositional literal analysis and direct figurative meaning retrieval when an idiom is encountered. However, whether and to what degree literal and figurative meanings are activated and used during comprehension depends on the relatedness of the idiom's meanings as well as how likely an idiom is to be idiomatic in a particular context. Unlike the Configuration Hypothesis, under the Hybrid Model activation of an idiom's figurative meaning depends strongly on semantics rather than word co-occurrences. The Hybrid Model particularly predicts that the degree of semantic relatedness between an idiom's literal and figurative meaning should have notable effects on processing.

Several results in the present study are congruent with the predictions made by the Hybrid Model. As previously mentioned in the Experiment 1B Discussion, our observation of lower accuracy for literal targets following highly figuratively-dominant idioms is congruent with the predictions of the Hybrid Model: the idiomatic meanings of more figuratively-dominant idioms are activated quickly even without supportive context, resulting in errors when the target is related to the literal meaning. Likewise, in Experiment 2 we found faster reading times for idioms with related literal and figurative meanings, again consistent with the predictions of the Hybrid Model. Under this view, the concurrent activation of literal and figurative meanings facilitates processing when the meanings are related.

However, the Hybrid Model and the Configuration Hypothesis are not ultimately contradictory. It's possible that the recognition point of an idiom might change depending on the dominance of an idiom's figurative meaning: more figuratively-dominant idioms might have earlier recognition points than less figuratively-dominant idioms. Literal analysis might also continue after the recognition point has been reached and the idiomatic meaning activated, resulting in facilitated comprehension when literal and figurative meanings are related.

In sum, although this study was not designed to test predictions made by specific models of idiom comprehension, our results nonetheless were congruent with several different models. In particular, our results support the characterization of idioms as lexicalized "chunks", but also support simultaneous literal analysis of idiom meaning.

4.4 IDIOM REPRESENTATION

Thus far, we have primarily discussed the present results in terms of their implications for idiom comprehension and processing rather than for mental representation of idioms. However, much of the research on ambiguous word processing is interpreted in light of its implications for the representation of ambiguous words in the lexicon—for example, whether the senses of a polyseme share a "core meaning" (Nunberg, 1979) or an underspecified representation (Eddington & Tokowicz, 2015; Frisson, 2009). Although the results of the present study do not definitively demonstrate that idioms and ambiguous words are processed similarly during comprehension, it is worth considering the implications for views of idiom representation if such processing similarities could be experimentally found.

Idioms and ambiguous words are both ambiguous units with multiple meanings that must be selected between during comprehension. However, these multiple meanings may not be represented in entirely the same ways. In the case of ambiguous words, the comprehender must choose between multiple stored meanings or senses, whether those meanings are stored separately, as for homonyms, or together, as for polysemes. The representation of idiom meaning is potentially more complex. Under most models of idiom comprehension, the figurative meaning is stored in the lexicon and retrieved during comprehension. However, the literal meaning—at least for the kinds of idioms discussed in the present study—is usually thought to be compositionally derived, and therefore not mentally represented in the same way that the figurative meaning is. This results in a different style of ambiguity for idioms and ambiguous words: ambiguity in words deriving from having to choose between which represented meaning is appropriate, and ambiguity in idioms deriving from whether to retrieve a stored meaning or continue with compositional literal analysis.

However, research on literal multiword phrases suggests that some multiword phrases may be processed similarly to single words. In particular, Arnon and Snider (2010) found the same effects of frequency on processing literal multiword phrases as are commonly found in words. They used these results to argue that multiword phrases might be represented in the lexicon, perhaps as a function of their frequency. However, they also found a lack of a clear processing distinction between highly-frequent phrases and other phrases, which they interpreted as evidence for a lack of distinction between exclusively stored and exclusively computed forms, as well as evidence for similar representation and processing of language regardless of grain size. Under this view, if a configuration of words is used literally with high frequency, the literal meaning of that phrase might be stored in the lexicon, even if that configuration of words also has a stored figurative meaning.

When thought of in this way, some idioms may be ambiguous in the same way as ambiguous words, with ambiguity deriving from which stored meaning to choose in a particular context. This raises the same questions asked about representations of ambiguous words: are different meanings stored separately or together? If together, is the representation based on core features, or are the different senses underspecified? Under the system of language processing investigated in the current study, in which the same constructs have the same processing implications regardless of whether ambiguity is single-word or multiword, meaning relatedness should affect idiom representation: some idioms would have representations more like homonyms, with literal and figurative meanings represented separately, and some idioms would have representations more like polysemes, with unitary representations for both literal and figurative meanings, whether underspecified or based on “core meanings”.

Under an underspecification account, the representation includes all semantically related senses of a word with which the comprehender is familiar. Unless a particular sense is required by the context, activating the underspecified representation is sufficient for comprehension. This account is congruent with the lack of effects following the idiom region in Experiment 2 of the present study: because the postcritical regions were neutral, comprehenders did not need to commit to a particular interpretation of the idiom, and the underspecified representation activated upon presentation of the idiom was sufficient for comprehension. Additionally, Frisson and Pickering (1999, 2001) have found evidence supporting an underspecification account in metonyms, raising the possibility that support for underspecification could extend to other forms of figurative language. However, in idioms as well as ambiguous words it is difficult to define exactly what content is underspecified.

Under a "core meaning" account, a representation contains only the features that are shared between senses. For example the core meaning of *rabbit* might contain the features +animate, +farm animal, +furry, +hop, and +big ears (Klepousniotou, 2008). Although some evidence has been found supporting this kind of representation, it has received criticism because it is difficult to determine what features could be shared between dominant and subordinate senses of a word: of the features listed, only +farm animal is completely compatible with the subordinate "meat" sense of *rabbit* (for an overview of this issue, see Brocher et al., 2016). It's likewise difficult to see how the literal and figurative meanings of an idiom could share enough features for a core meaning account to be plausible.

The present work, as well as the theories of representation discussed above, highlights the question of what idioms are. One possibility is that idiom meanings are abstractions computed over world knowledge, similar to the conceptualization of selectional restrictions proposed by Warren and colleagues (2015). Under this view, the meaning of an idiom is developed based on the experience an individual comprehender has with both the linguistic contexts in which an idiom appears and the events it is likely to describe. The high variability in idiom meaning seen between individual speakers and across different locations points to a strong role for experience in creation of idiom meaning (Nordmann, Clelland, & Bull, 2014; Milburn, unpublished norms). Additionally, Hamblin and Gibbs (1999) found that characteristics of single words within an idiom limited the events that idiom could appropriately describe (for example, *kick the bucket* could describe quick deaths but not slow ones). This suggests that comprehenders are able to generalize appropriateness of idioms to new situations based on their world knowledge about what events idioms are usually used to describe. Although more research is needed to investigate this possibility empirically, evidence exists for a strong role of world knowledge in idiom comprehension, leading to the possibility that world knowledge as well as the lexical-level constructs discussed elsewhere in this document interact to both create idiom representations and drive idiom processing.

4.5 CONCLUSION

The present study provides preliminary evidence that idioms and ambiguous words are treated similarly during ambiguity resolution. Although our comparisons between idioms and ambiguous words were not entirely conclusive, we found some similarities between processing of idioms and ambiguous words, particularly in eyetracking during reading. In particular, we found that idioms with highly figuratively-dominant meanings and high relatedness between meanings showed a processing disadvantage, similar to effects found in ambiguous words. Additionally, we replicated the polysemy advantage observed in ambiguous word research in idioms, suggesting that the construct of meaning relatedness is universal to ambiguity resolution regardless of grain size or literality. These results have implications for our understanding of idiom comprehension, and suggest valuable new avenues for future research.

APPENDIX A

MODEL RESULTS

A.1 EXPERIMENT 1A

Table 17. Model estimates for Experiment 1A priming effects on reaction time

Effect	Estimate	SE	t-value	p-value
Intercept	**-0.07**	**0.02**	**-3.15**	**0.00**
Prime Type	**-0.02**	**0.01**	**-2.53**	**0.01**
Target Type	0.00	0.01	-0.23	0.82
Nonword Bigram Frequency	0.01	0.04	0.21	0.84
Trial Number	**0.00**	**0.00**	**3.97**	**0.00**
Previous Trial RT	**0.14**	**0.02**	**8.74**	**0.00**
Target Word Bigram Frequency	0.00	0.01	0.01	1.00
Target Word Length	-0.01	0.01	-1.24	0.22
Target Word # Syllables	**0.04**	**0.01**	**4.97**	**0.00**
Target Word Concreteness	-0.01	0.01	-1.20	0.23
Previous Trial Accuracy	*-0.03*	*0.02*	*-1.83*	*0.07*

Prime Type*Target Type	0.01	0.02	0.50	0.61
Prime Type*Nonword Bigram Frequency	0.02	0.02	1.17	0.24
Target Type* Nonword Bigram Frequency	0.01	0.02	0.36	0.72
Prime Type * Target Type * Nonword Bigram Frequency	0.00	0.03	-0.06	0.95

Table 18. Model estimates for Experiment 1A dominance/relatedness effects on reaction time

Effect	Estimate	SE	t-value	p-value
Intercept	**-0.06**	**0.03**	**-2.10**	**0.04**
Meaning Dominance (MD)	**-0.01**	**0.01**	**-2.13**	**0.03**
Ambiguity Type	0.00	0.02	-0.33	0.75
Target Type	0.01	0.01	0.84	0.41
Nonword Bigram Frequency	0.01	0.04	0.21	0.83
Prime-Target Relatedness Score	-0.02	0.02	-0.94	0.35
Trial Number	**0.00**	**0.00**	**3.43**	**0.00**
Previous Trial RT	**0.19**	**0.02**	**8.91**	**0.00**
Target Word Bigram Frequency	0.00	0.01	0.42	0.68
Target Word Length	*-0.02*	*0.01*	*-1.81*	*0.07*
Target Word # Syllables	**0.04**	**0.01**	**3.97**	**0.00**
Target Word Concreteness	-0.01	0.01	-1.50	0.13
Previous Trial Accuracy	**-0.05**	**0.02**	**-2.13**	**0.03**
Meaning Dominance* Ambiguity Type	0.00	0.01	0.09	0.93

MD*Target Type	-0.02	0.01	-1.40	0.17
Ambiguity Type *Target Type	0.01	0.02	0.56	0.57
MD* Nonword Bigram Frequency	0.01	0.01	1.31	0.19
Ambiguity Type * Nonword Bigram Frequency	0.01	0.02	0.48	0.63
Target Type* Nonword Bigram Frequency	-0.02	0.02	-0.92	0.36
MD Ambiguity Type *Target Type*	*-0.05*	*0.03*	*-1.79*	*0.08*
MD* Ambiguity Type * Nonword Bigram Frequency	**0.04**	**0.02**	**1.98**	**0.05**
MD Target Type* Nonword Bigram Frequency*	*0.04*	*0.02*	*1.83*	*0.07*
Ambiguity Type * Target Type* Nonword Bigram Frequency	-0.03	0.04	-0.65	0.52
MD* Ambiguity Type * Target Type* Nonword Bigram Frequency	-0.01	0.04	-0.13	0.89

Table 19. Model estimates for Experiment 1A effects on reaction time within homonyms and polysemes

Ambiguity Type	Effect	Estimate	SE	t-value	p-value
Homonyms	Intercept	-0.05	0.03	-1.42	0.16
	Meaning Dominance (MD)	**-0.01**	**0.01**	**-1.97**	**0.05**
	Nonword Bigram Frequency	0.00	0.03	0.09	0.93
	Prime-Target Relatedness Score	0.00	0.02	-0.14	0.89
	Trial Number	**0.00**	**0.00**	**2.48**	**0.01**
	Previous Trial RT	**0.22**	**0.03**	**7.04**	**0.00**
	Target Word Bigram Frequency	-0.01	0.01	-0.81	0.42
	Target Word Length	**-0.04**	**0.02**	**-2.55**	**0.01**

	Target Word # Syllables		**0.06**	**0.02**	**4.15**	**0.00**
	Target Word Concreteness		-0.01	0.01	-0.59	0.56
	Previous Trial Accuracy		**-0.08**	**0.03**	**-2.59**	**0.01**
	Meaning Dominance* Nonword Bigram Frequency		-0.01	0.01	-0.43	0.67
Polysemes	*Intercept*		*-0.07*	*0.04*	*-1.77*	*0.08*
	Meaning Dominance (MD)		-0.01	0.01	-1.07	0.29
	Nonword Bigram Frequency		0.01	0.04	0.35	0.73
	Prime-Target Relatedness Score		-0.03	0.03	-1.26	0.21
	Trial Number		**0.00**	**0.00**	**2.25**	**0.02**
	Previous Trial RT		**0.20**	**0.03**	**6.78**	**0.00**
	Target Word Bigram Frequency		0.01	0.01	1.02	0.31
	Target Word Length		0.00	0.02	-0.04	0.97
	Target Word # Syllables		0.02	0.02	1.54	0.13
	Target Word Concreteness		-0.01	0.01	-0.85	0.40
	Previous Trial Accuracy		-0.01	0.03	-0.21	0.83
	Meaning Dominance* Nonword Bigram Frequency		**0.03**	**0.02**	**2.13**	**0.03**
	Higher Nonword Bigram Frequency	Intercept	-0.05	0.06	-0.80	0.42
		Meaning Dominance (MD)	*-0.02*	*0.01*	*-1.88*	*0.06*
		Prime-Target Relatedness Score	*-0.07*	*0.04*	*-1.94*	*0.054*
		Trial Number	0.00	0.00	1.41	0.16
		Previous Trial RT	**0.16**	**0.04**	**3.92**	**0.00**

	Target Word Bigram Frequency	0.03	0.02	1.64	0.11
	Target Word Length	0.01	0.02	0.58	0.56
	Target Word # Syllables	0.01	0.02	0.67	0.50
	Target Word Concreteness	0.01	0.01	0.70	0.48
	Previous Trial Accuracy	0.01	0.05	0.13	0.89
Lower Nonword Bigram Frequency	Intercept	-0.07	0.05	-1.26	0.21
	Meaning Dominance (MD)	0.01	0.01	0.80	0.42
	Prime-Target Relatedness Score	-0.03	0.03	-0.75	0.45
	Trial Number	0.00	0.00	1.29	0.20
	Previous Trial RT	**0.26**	**0.04**	**5.81**	**0.00**
	Target Word Bigram Frequency	-0.01	0.02	-0.37	0.71
	Target Word Length	-0.01	0.02	-0.76	0.45
	Target Word # Syllables	*0.04*	*0.02*	*1.88*	*0.06*
	Target Word Concreteness	**-0.02**	**0.01**	**-2.05**	**0.04**
	Previous Trial Accuracy	0.00	0.05	-0.08	0.94

Table 20. Model estimates for Experiment 1A priming effects on accuracy

Effect	Estimate	SE	t-value	p-value
Intercept	1.95	0.07	26.57	.00
Nonword Bigram Frequency	-0.01	0.01	-0.93	0.35
Prime Type	**0.24**	**0.08**	**3.16**	**0.00**

Target Type	0.089	0.08	1.13	0.26
Target Word Bigram Frequency	**-0.12**	**0.052**	**-2.23**	**0.03**
Target Word Length	**0.29**	**0.07**	**4.18**	**.00**
Target Word # of Syllables	**-0.15**	**0.07**	**-2.03**	**0.04**
Target Word Concreteness	0.02	0.04	.00	0.59
Nonword Bigram Frequency*Prime Type	0.01	0.02	0.46	0.65
Nonword Bigram Frequency*Target Type	0.03	0.02	1.57	0.12
Prime Type*Target Type	0.07	0.15	0.46	0.65
Nonword Bigram Frequency* Prime Type*Target Type	-0.03	0.04	-0.72	0.47

Table 21. Model estimates for Experiment 1A MD/MR effects on accuracy

Effect	Estimate	SE	t-value	p-value
Intercept	**1.89**	**0.14**	**13.58**	**0.00**
Nonword Bigram Frequency	-0.04	0.08	-0.57	0.57
Target Type	0.00	0.08	0.04	0.97
Ambiguity Type	0.05	0.14	0.33	0.74
Meaning Dominance	0.05	0.04	1.30	0.20
Prime-Target Relatedness Score	*0.23*	*0.13*	*1.80*	*0.07*
Target Word Bigram Frequency	**-0.15**	**0.07**	**-2.20**	**0.03**
Target Word Length	**0.30**	**0.09**	**3.51**	**0.00**
Target Word # of Syllables	-0.12	0.09	-1.34	0.18

Target Word Concreteness	-0.02	0.05	-0.42	0.68
Nonword Bigram Frequency*Target Type	0.08	0.15	0.52	0.61
*Nonword Bigram Frequency*Ambiguity Type*	*-0.26*	*0.15*	*-1.71*	*0.09*
*Target Type*Ambiguity Type*	*-0.31*	*0.16*	*-1.94*	*0.053*
Nonword Bigram Frequency*MD	0.08	0.08	1.00	0.32
Target Type*MD	0.04	0.14	0.28	0.78
*Ambiguity Type*MD*	*-0.15*	*0.08*	*-1.86*	*0.06*
Nonword Bigram Frequency*Target Type*Ambiguity Type	-0.24	0.31	-0.79	0.43
Nonword Bigram Frequency*Target Type*MD	0.15	0.16	0.97	0.33
Nonword Bigram Frequency*Ambiguity Type*MD	-0.06	0.16	-0.37	0.71
Target Type*Ambiguity Type*MD	0.35	0.29	1.22	0.23
Nonword Bigram Frequency*Target Type*Ambiguity Type*MD	0.22	0.32	0.68	0.49

Table 22. Model estimates for Experiment 1A MD effects on accuracy within each ambiguity type

Ambiguity Type	Effect	Estimate	SE	t-value	p-value
Homonyms	**Intercept**	**1.86**	**0.18**	**10.29**	**0.00**
	Nonword Bigram Frequency	0.09	0.11	0.83	0.41
	Target Type	0.13	0.12	1.08	0.28
	Meaning Dominance	*0.11*	*0.06*	*1.87*	*0.06*
	Prime-Target Relatedness Score	0.15	0.16	0.93	0.35
	Target Word Bigram Frequency	**-0.31**	**0.09**	**-3.58**	**0.00**

	Target Word Length	**0.51**	**0.13**	**3.95**	**0.00**
	Target Word # of Syllables	-0.22	0.14	-1.66	0.10
	Target Word Concreteness	0.03	0.08	0.43	0.67
	Nonword Bigram Frequency*Target Type	0.19	0.22	0.87	0.39
	Nonword Bigram Frequency*MD	0.11	0.11	1.00	0.32
	Target Type*MD	-0.08	0.19	-0.44	0.66
	Nonword Bigram Frequency*Target Type*MD	0.06	0.23	0.25	0.81
Polysemes	**Intercept**	**1.89**	**0.22**	**8.55**	**0.00**
	Nonword Bigram Frequency	*-0.18*	*0.10*	*-1.73*	*0.09*
	Target Type	-0.15	0.12	-1.23	0.22
	Meaning Dominance	-0.03	0.06	-0.53	0.60
	Prime-Target Relatedness Score	0.35	0.21	1.65	0.10
	Target Word Bigram Frequency	0.08	0.10	0.78	0.44
	Target Word Length	0.09	0.11	0.78	0.44
	Target Word # of Syllables	-0.03	0.12	-0.23	0.82
	Target Word Concreteness	-0.07	0.08	-0.84	0.40
	Nonword Bigram Frequency*Target Type	-0.05	0.21	-0.25	0.80
	Nonword Bigram Frequency*MD	0.05	0.11	0.44	0.66
	Target Type*MD	0.15	0.20	0.73	0.47
	Nonword Bigram Frequency*Target Type*MD	0.25	0.22	1.14	0.26

A.2 EXPERIMENT 1B

Table 23. Model estimates for Experiment 1B priming effects on reaction time

Effect	Estimate	SE	t-value	p-value
Intercept	*0.16*	*0.09*	*1.81*	*0.07*
Prime Type	**-0.13**	**0.04**	**-3.42**	**0.00**
Target Type	**0.14**	**0.07**	**2.07**	**0.04**
Nonword Bigram Frequency	0.07	0.13	0.56	0.58
Trial Number	0.00	0.00	-0.62	0.54
Previous Trial Reaction Time	**0.14**	**0.02**	**7.10**	**0.00**
Target Word Bigram Frequency	**0.07**	**0.03**	**2.29**	**0.02**
Target Word Length	**-0.14**	**0.04**	**-3.99**	**0.00**
Target Word # of Syllables	0.03	0.04	0.65	0.51
Target Word Concreteness	**0.10**	**0.04**	**2.57**	**0.01**
Previous Trial Accuracy	0.12	0.08	1.62	0.11
Prime Type*Target Type	**0.16**	**0.07**	**2.41**	**0.02**
Prime Type* Nonword Bigram Frequency	0.06	0.07	0.83	0.41
Target Type* Nonword Bigram Frequency	0.07	0.07	1.07	0.28
Prime Type*Target Type* Nonword Bigram Frequency	0.07	0.13	0.55	0.59

Table 24. Model estimates for Experiment 1B priming effects within each prime type on reaction time

Prime Type	Effect	Estimate	SE	t-value	p-value
Idiom	**Intercept**	**0.23**	**0.10**	**2.44**	**0.02**
	Target Type	0.09	0.08	1.08	0.28
	Nonword Bigram Frequency	0.05	0.13	0.36	0.72
	Trial Number	0.00	0.00	-0.33	0.74
	Previous Trial Reaction Time	**0.18**	**0.03**	**6.87**	**0.00**
	Target Word Bigram Frequency	0.06	0.04	1.57	0.12
	Target Word Length	**-0.13**	**0.05**	**-2.95**	**0.00**
	Target Word # of Syllables	0.03	0.05	0.64	0.53
	Target Word Concreteness	**0.12**	**0.05**	**2.53**	**0.01**
	Previous Trial Accuracy	0.10	0.11	0.96	0.34
	Target Type* Nonword Bigram Frequency	0.05	0.09	0.48	0.63
Control	Intercept	0.09	0.11	0.83	0.41
	Target Type	*0.27*	*0.14*	*1.89*	*0.06*
	Nonword Bigram Frequency	0.11	0.14	0.75	0.46
	Trial Number	0.00	0.00	-0.59	0.56
	Previous Trial Reaction Time	**0.13**	**0.03**	**4.60**	**0.00**
	Target Word Bigram Frequency	0.05	0.06	0.84	0.40

Target Word Length	*-0.13*	*0.07*	*-1.94*	*0.06*
Target Word # of Syllables	-0.02	0.07	-0.31	0.76
Target Word Concreteness	0.13	0.08	1.69	0.10
Previous Trial Accuracy	0.16	0.11	1.44	0.15
Target Type* Nonword Bigram Frequency	0.10	0.09	1.12	0.26

Table 25. Model estimates for Experiment 1B dominance/relatedness effects on reaction time

Effect	Estimate	SE	t-value	p-value
Intercept	**0.28**	**0.09**	**2.97**	**0.00**
Meaning Dominance	-0.03	0.06	-0.49	0.63
Meaning Relatedness	-0.06	0.04	-1.55	0.13
Nonword Bigram Frequency	0.04	0.13	0.33	0.74
Target Type	0.11	0.09	1.21	0.23
Trial Number	0.00	0.00	-0.44	0.66
Previous Trial Reaction Time	**0.17**	**0.03**	**6.60**	**0.00**
Target Word Bigram Frequency	0.05	0.04	1.34	0.18
Target Word Length	**-0.10**	**0.05**	**-2.30**	**0.02**
Target Word # of Syllables	**0.00**	**0.05**	**0.00**	**1.00**
Target Word Concreteness	**0.12**	**0.05**	**2.54**	**0.01**
Previous Trial Accuracy	0.11	0.11	1.05	0.29
Idiom Familiarity	0.07	0.05	1.50	0.14

MD*MR	-0.01	0.05	-0.14	0.89
Nonword Bigram Frequency*MD	0.04	0.05	0.81	0.42
Nonword Bigram Frequency*MR	-0.06	0.05	-1.22	0.22
MD*Target Type	0.07	0.06	1.22	0.22
MR*Target Type	-0.01	0.05	-0.20	0.84
Nonword Bigram Frequency*Target Type	0.00	0.10	-0.04	0.97
MD*MR* Nonword Bigram Frequency	0.00	0.06	-0.01	0.99
MD*MR*Target Type	0.05	0.06	0.81	0.42
MD* Nonword Bigram Frequency*Target Type	0.06	0.11	0.54	0.59
MR* Nonword Bigram Frequency*Target Type	0.01	0.10	0.13	0.90
MD* MR* Nonword Bigram Frequency*Target Type	0.10	0.11	0.85	0.40

Table 26. Model estimates for Experiment 1B priming effects on accuracy

Effect	Estimate	SE	t-value	p-value
Intercept	**2.18**	**0.08**	**27.89**	**0.00**
Nonword Bigram Frequency	**0.05**	**0.01**	**3.95**	**0.00**
Prime Type	0.18	0.14	1.25	0.21
Target Type	0.14	0.08	1.62	0.11
Target Word Bigram Frequency	*0.09*	*0.05*	*1.74*	*0.08*
Target Word Length	*0.13*	*0.07*	*1.96*	*0.051*
Target Word # of Syllables	**-0.15**	**0.07**	**-1.97**	**0.05**

Target Word Concreteness	0.06	0.07	0.78	0.43
Nonword Bigram Frequency*Target Type	**0.05**	**0.02**	**2.04**	**0.04**
Nonword Bigram Frequency*Prime Type	-0.04	0.02	-1.54	0.12
Prime Type*Target Type	0.16	0.17	0.95	0.34
Nonword Bigram Frequency* Prime Type*Target Type	-0.01	0.05	-0.21	0.84

Table 27. Model estimates for Experiment 1B accuracy effects within each level of nonword bigram frequency

Nonword Bigram Frequency	Effect	Estimate	SE	t-value	p-value
High Frequency	**Intercept**	**1.97**	**0.09**	**22.07**	**0.00**
	Target Type	-0.01	0.19	-0.06	0.96
	Prime Type	**0.32**	**0.10**	**3.27**	**0.00**
	Target Word Bigram Frequency	*0.13*	*0.08*	*1.74*	*0.08*
	Target Word Length	0.08	0.10	0.81	0.42
	Target Word # Syllables	-0.08	0.11	-0.71	0.48
	Target Word Concreteness	0.10	0.11	0.92	0.36
	Target Type*Prime Type	0.23	0.19	1.19	0.24
Low Bigram Frequency	**Intercept**	**2.21**	**0.07**	**30.86**	**0.00**
	Target Type	0.21	0.15	1.47	0.15
	Prime Type	0.11	0.08	1.52	0.13
	Target Word Bigram Frequency	0.06	0.06	1.01	0.32

Target Word Length	**0.18**	**0.07**	**2.45**	**0.02**
Target Word # Syllables	**-0.19**	**0.08**	**-2.33**	**0.02**
Target Word Concreteness	0.10	0.08	1.22	0.22
Target Type*Prime Type	0.16	0.15	1.06	0.29

Table 28. Model estimates for Experiment 1B dominance/relatedness effects on accuracy

Effect	Estimate	SE	t-value	p-value
Intercept	**2.20**	**0.09**	**24.48**	**0.00**
Nonword Bigram Frequency	0.01	0.02	0.91	0.36
Target Type	0.07	0.18	0.38	0.71
Meaning Dominance	-0.06	0.10	-0.60	0.55
Meaning Relatedness	-0.10	0.09	-1.13	0.26
Idiom Familiarity	0.02	0.09	0.23	0.82
Target Word Bigram Frequency	0.10	0.07	1.36	0.18
Target Word Length	0.06	0.09	0.70	0.48
Target Word # of Syllables	-0.07	0.10	-0.68	0.50
Target Word Concreteness	-0.03	0.09	-0.28	0.78
Nonword Bigram Frequency*Target Type	0.03	0.03	1.03	0.31
Nonword Bigram Frequency*MD	-0.01	0.02	-0.58	0.56
Target Type*MD	**0.31**	**0.14**	**2.24**	**0.03**
Nonword Bigram Frequency*MR	-0.02	0.02	-0.94	0.35
Target Type*MR	0.04	0.13	0.30	0.76

MD*MR	0.07	0.10	0.65	0.52
Nonword Bigram Frequency * Target Type * MD	0.06	0.04	1.68	0.10
*Nonword Bigram Frequency * Target Type*MR*	*0.06*	*0.03*	*1.90*	*0.06*
Nonword Bigram Frequency * MD * MR	0.01	0.02	0.29	0.77
Target Type*MD*MR	0.05	0.14	0.35	0.73
Nonword Bigram Frequency * Target Type * MD*MR	0.01	0.04	0.20	0.84

A.3 EXPERIMENT 2

Table 29. Model estimates for Experiment 2 eye tracking measures in the precritical region

Measure	Effect	Estimate	SE	t-value	p-value
First Fixation	*Intercept*	*143.54*	*79.04*	*40.50*	*0.08*
	Context Bias	103.41	136.02	1475.60	0.45
	Meaning Dominance	4.33	10.80	39.30	0.69
	Meaning Relatedness	7.64	20.86	40.50	0.72
	Familiarity	6.89	4.89	42.00	0.17
	Progressive Naturalness Score	-0.72	3.49	75.10	0.84
	Context Bias*MD	-14.63	19.08	1471.90	0.44
	Context Bias*MR	-16.38	37.05	1473.30	0.66
	MD*MR	-0.92	2.89	39.00	0.75
	Context Bias*MD*MR	2.29	5.12	1471.10	0.66
Go Past	Intercept	819.13	729.37	1.12	>.05

	Context Bias	-260.68	272.47	-0.96	>.05
	Meaning Dominance	-30.04	79.61	-0.38	>.05
	Meaning Relatedness	-84.54	207.53	-0.41	>.05
	Familiarity	-22.35	62.02	-0.36	>.05
	Progressive Naturalness Score	-7.27	15.42	-0.47	>.05
	Context Bias*MD	46.15	35.95	1.28	>.05
	Context Bias*MR	100.35	94.63	1.06	>.05
	MD*MR	11.73	27.88	0.42	>.05
	Context Bias*MD*MR	-16.18	12.55	-1.29	>.05
Re-Reading	Intercept	549.51	528.11	1.04	0.30
	Context Bias	-252.20	725.49	-0.35	0.73
	Meaning Dominance	-46.13	73.30	-0.63	0.53
	Meaning Relatedness	-69.07	140.77	-0.49	0.63
	Familiarity	3.69	33.50	0.11	0.91
	Progressive Naturalness Score	16.24	22.62	0.72	0.47
	Context Bias*MD	47.48	102.68	0.46	0.64
	Context Bias*MR	84.52	198.72	0.43	0.67
	MD*MR	10.14	19.70	0.52	0.61
	Context Bias*MD*MR	-15.03	27.58	-0.55	0.59
Total Time	Intercept	1104.20	1261.77	0.88	0.39
	Context Bias	-279.36	490.16	-0.57	0.57

Meaning Dominance	-76.59	173.59	-0.44	0.66
Meaning Relatedness	-174.49	335.72	-0.52	0.61
Familiarity	-22.20	77.67	-0.29	0.78
Progressive Naturalness Score	29.53	16.42	1.80	0.10
Context Bias*MD	56.49	68.57	0.82	0.41
Context Bias*MR	102.21	132.50	0.77	0.44
MD*MR	20.88	46.36	0.45	0.65
Context Bias*MD*MR	-18.70	18.24	-1.03	0.31

Table 30. Model estimates for Experiment 2 eye tracking measures in the critical region

Measure	Effect	Estimate	SE	t-value	p-value
First Fixation	**Intercept**	**212.27**	**88.78**	**2.39**	**0.02**
	Context Bias	-36.33	110.86	-0.33	0.74
	Meaning Dominance	-1.41	12.07	-0.12	0.91
	Meaning Relatedness	-7.69	23.37	-0.33	0.74
	Familiarity	2.04	5.44	0.37	0.71
	Progressive Naturalness Score	0.47	2.65	0.18	0.86
	Context Bias*MD	1.73	15.53	0.11	0.91
	Context Bias*MR	11.77	30.25	0.39	0.70
	MD*MR	0.97	3.23	0.30	0.77
	Context Bias*MD*MR	-0.80	4.17	-0.19	0.85
Go Past	**Intercept**	**2695.34**	**646.22**	**4.17**	**0.00**

	Context Bias	416.35	437.45	0.95	0.34
	Meaning Dominance	**-254.06**	**87.97**	**-2.89**	**0.01**
	Meaning Relatedness	**-547.01**	**170.21**	**-3.21**	**0.00**
	Familiarity	*-71.54*	*39.35*	*-1.82*	*0.08*
	Progressive Naturalness Score	7.77	13.28	0.59	0.56
	Context Bias*MD	-42.06	61.27	-0.69	0.49
	Context Bias*MR	-91.84	119.73	-0.77	0.44
	MD*MR	**76.96**	**23.50**	**3.28**	**0.00**
	Context Bias*MD*MR	8.67	16.48	0.53	0.60
Re-Reading	**Intercept**	**1217.47**	**328.66**	**3.70**	**0.00**
	Context Bias	673.64	603.51	1.12	0.26
	Meaning Dominance	*-91.86*	*45.61*	*-2.01*	*0.051*
	Meaning Relatedness	**-197.38**	**87.89**	**-2.25**	**0.03**
	Familiarity	*-38.93*	*20.33*	*-1.92*	*0.06*
	Progressive Naturalness Score	-1.89	11.51	-0.16	0.87
	Context Bias*MD	-92.52	86.00	-1.08	0.28
	Context Bias*MR	-148.51	166.01	-0.90	0.37
	MD*MR	**26.60**	**12.24**	**2.17**	**0.04**
	Context Bias*MD*MR	20.61	23.13	0.89	0.37
Total Time	**Intercept**	**2646.10**	**668.72**	**3.96**	**0.00**
	Context Bias	156.55	492.17	0.32	0.75

Meaning Dominance	**-214.91**	**91.01**	**-2.36**	**0.02**
Meaning Relatedness	**-465.92**	**176.08**	**-2.65**	**0.01**
Familiarity	**-91.72**	**40.71**	**-2.25**	**0.03**
Progressive Naturalness Score	-4.21	14.91	-0.28	0.78
Context Bias*MD	-6.61	68.95	-0.10	0.92
Context Bias*MR	28.34	134.57	0.21	0.83
MD*MR	**65.65**	**24.31**	**2.70**	**0.01**
Context Bias*MD*MR	-7.82	18.53	-0.42	0.67

Table 31. Model estimates for Experiment 2 eye tracking measures in the postcritical region

Measure	Effect	Estimate	SE	t-value	p-value
First Fixation	*Intercept*	*156.04*	*83.41*	*1.87*	*0.07*
	Context Bias	-195.79	134.23	-1.46	0.14
	Meaning Dominance	15.69	11.48	1.37	0.18
	Meaning Relatedness	22.12	22.12	1.00	0.32
	Familiarity	-6.66	5.09	-1.31	0.20
	Progressive Naturalness Score	-3.25	3.77	-0.86	0.39
	*Context Bias*MD*	*35.29*	*18.81*	*1.88*	*0.06*
	Context Bias*MR	50.60	36.39	1.39	0.16
	MD*MR	-2.77	3.06	-0.91	0.37
	*Context Bias*MD*MR*	*-9.06*	*5.02*	*-1.81*	*0.07*
	Literal Context **Intercept**	**240.84**	**107.93**	**2.23**	**0.03**

		Meaning Dominance	-3.15	14.64	-0.22	0.83
		Meaning Relatedness	-6.61	28.40	-0.23	0.82
		Idiom Familiarity	-5.73	6.54	-0.88	0.39
		Progressive Naturalness Score	0.09	5.09	0.02	0.99
		MD*MR	2.27	3.92	0.58	0.57
	Figurative Context	**Intercept**	70.65	108.33	0.65	0.52
		Meaning Dominance	**34.96**	**15.28**	**2.29**	**0.03**
		Meaning Relatedness	48.89	29.28	1.67	0.10
		Idiom Familiarity	-8.76	6.78	-1.29	0.20
		Progressive Naturalness Score	-6.25	5.75	-1.09	0.28
		*MD*MR*	*-7.50*	*4.07*	*-1.84*	*0.07*
Go Past	Intercept		1377.14	979.23	1.41	0.17
	Context Bias		177.49	769.47	0.23	0.82
	Meaning Dominance		-41.43	133.82	-0.31	0.76
	Meaning Relatedness		-47.56	258.63	-0.18	0.86
	Familiarity		5.83	60.49	0.10	0.92
	Progressive Naturalness Score		**-74.63**	**25.38**	**-2.94**	**0.00**
	Context Bias*MD		-12.89	107.80	-0.12	0.90
	Context Bias*MR		71.30	208.31	0.34	0.73
	MD*MR		7.31	35.74	0.21	0.84
	Context Bias*MD*MR		-13.55	28.72	-0.47	0.64
Second Pass	Intercept		-259.45	739.60	-0.35	0.73

	Context Bias	-222.97	1315.16	-0.17	0.87
	Meaning Dominance	82.18	102.94	0.80	0.43
	Meaning Relatedness	125.48	193.47	0.65	0.52
	Familiarity	13.85	43.85	0.32	0.75
	Progressive Naturalness Score	18.86	35.53	0.53	0.60
	Context Bias*MD	34.68	188.71	0.18	0.85
	Context Bias*MR	82.32	351.17	0.23	0.82
	MD*MR	-21.05	27.17	-0.78	0.44
	Context Bias*MD*MR	-13.78	49.53	-0.28	0.78
Total Time	Intercept	1346.53	987.15	1.36	0.18
	Context Bias	507.86	518.71	0.98	0.33
	Meaning Dominance	-45.18	134.67	-0.34	0.74
	Meaning Relatedness	-75.01	260.30	-0.29	0.77
	Familiarity	-33.90	60.12	-0.56	0.58
	Progressive Naturalness Score	**-39.47**	**17.62**	**-2.24**	**0.03**
	Context Bias*MD	-58.94	72.68	-0.81	0.42
	Context Bias*MR	-23.15	140.15	-0.17	0.87
	MD*MR	12.15	35.96	0.34	0.74
	Context Bias*MD*MR	0.02	19.34	0.00	1.00

APPENDIX B

STIMULI

B.1 EXPERIMENT 1A

Table 32. Mean concreteness, frequency, dominance score, and prime-target relatedness scores for Experiment 1A ambiguous word stimuli

Ambiguous Prime	Control Prime	Dominant Target	Subordinate Target	Ambiguity Type	Ambig. Conc. / Cont. Conc.	Ambig. Freq. / Cont. Freq.	Dom. Conc. / Sub. Conc.	Dom. Freq. / Sub. Freq.	Dom. Score / Sub. Score	MR Ambig /Dom. / Ambig /Sub.	MR Cont/ Dom. / Cont/ Sub
arena	scalp	stadium	scene	polyseme	4.83	2.27	4.79	2.50	51.28	1.17	6.13
					4.82	2.28	3.93	3.58	48.72	3.63	6.00
article	trigger	newspaper	item	polyseme	4.33	3.00	4.82	3.08	54.60	1.40	5.67
					4.31	2.94	4.41	2.80	45.40	2.33	5.25
atmosphere	commission	air	mood	polyseme	3.04	2.69	4.11	3.85	55.40	2.00	5.83
					3.04	2.85	1.75	3.24	44.60	1.40	5.50
bar	van	cocktail	soap	homonym	4.67	3.64	4.40	2.75	53.68	1.67	5.25
					4.72	3.42	4.93	2.89	46.32	1.63	5.57

bark	grid	yelp	rind	homonym	4.52	2.45	3.54	0.90	56.32	2.00	6.00
					4.55	2.45	4.48	1.18	43.68	4.13	6.00
bat	fox	vampire	mitt	homonym	5.00	3.02	4.18	2.95	56.52	1.83	5.71
					4.97	3.04	4.76	1.85	43.48	1.63	5.00
beam	chew	light	wood	polyseme	3.96	2.65	4.21	3.93	56.56	1.88	5.00
					3.93	2.67	4.85	3.14	43.44	2.00	4.43
blade	shark	knife	grass	polyseme	4.93	2.82	4.90	3.38	50.60	1.17	5.25
					4.93	2.88	4.93	2.93	49.40	2.40	6.33
bolt	rash	lightning	wrench	homonym	4.67	2.55	4.59	2.86	54.76	1.40	6.75
					4.62	2.61	4.93	2.31	45.24	1.94	5.89
border	hockey	frame	patrol	polyseme	4.29	2.94	4.30	2.86	54.48	2.83	5.33
					4.31	2.77	3.86	2.90	45.52	2.88	6.00
bottle	dollar	nipple	flask	polyseme	4.91	3.41	4.83	2.16	51.56	2.75	5.75
					4.93	3.15	4.79	1.76	48.44	2.00	6.00
calf	snot	bull	shin	homonym	4.48	2.18	4.85	3.15	52.16	1.88	4.75
					4.48	2.04	4.86	2.20	47.84	1.00	5.71
cape	tray	cloak	peninsula	homonym	4.77	2.62	4.71	2.19	54.12	2.17	6.29
					4.74	2.61	4.86	1.85	45.88	2.25	6.75
chain	grave	link	series	polyseme	4.55	3.03	3.43	2.79	57.40	2.00	6.75
					4.56	3.13	2.92	3.01	42.60	2.00	5.17
check	write	cross	debt	homonym	4.11	4.15	4.44	3.45	53.52	5.00	4.29
					4.22	3.81	2.72	2.86	46.48	2.50	6.25
coat	bell	sweater	hair	polyseme	4.97	3.33	4.78	2.85	62.72	2.13	6.00
					4.96	3.30	4.97	3.89	37.28	3.00	6.50
cone	cork	geometry	sorbet	polyseme	4.86	2.18	3.00	1.98	52.60	2.00	5.57
					4.86	2.17	4.43	1.18	47.40	3.19	5.33
cotton	bucket	wool	plant	polyseme	4.97	2.86	4.86	2.21	52.76	2.00	5.67
					4.96	2.71	4.76	3.15	47.24	1.88	4.00
deck	beef	balcony	card	homonym	4.77	3.08	4.68	2.57	54.28	1.63	5.57
					4.74	3.00	4.90	3.64	45.72	1.67	7.00
degree	visual	scale	doctorate	polyseme	3.00	2.88	4.39	2.69	51.20	2.67	4.00
					2.96	2.77	3.40	1.67	48.80	2.60	6.13
doll	soda	nice	toy	polyseme	5.00	3.10	2.18	4.52	50.92	2.83	6.88

					4.97	3.01	4.93	2.93	49.08	1.63	5.86
drill	blond	tool	practice	homonym	4.40	2.85	4.60	2.74	53.28	1.17	6.38
					4.41	2.73	2.52	3.37	46.72	3.40	5.67
gas	fly	vapor	oil	polyseme	4.29	3.54	4.15	1.51	55.08	1.83	5.00
					4.64	3.64	4.93	3.32	44.92	1.80	5.71
gear	edge	equipment	motor	homonym	4.28	2.91	4.83	3.11	57.40	1.50	5.14
					4.24	3.08	4.84	2.83	42.60	2.50	6.50
gem	bun	jewel	masterpiece	polyseme	4.88	2.10	4.96	2.57	62.52	1.33	6.00
					4.88	2.17	3.11	2.26	37.48	2.88	5.29
glass	penny	window	cup	polyseme	4.82	3.49	4.86	3.64	52.88	1.80	5.78
					4.83	3.09	5.00	3.42	47.12	1.83	5.81
goal	tone	dream	score	polyseme	3.06	2.93	2.60	3.83	55.96	1.50	5.50
					3.07	2.94	3.38	3.19	44.04	1.50	4.00
lap	hug	swimmer	chair	homonym	4.11	2.84	4.77	2.28	50.72	1.50	5.86
					4.14	2.99	4.58	3.40	49.28	3.00	5.75
litter	bomber	trash	puppy	homonym	4.47	2.30	4.70	3.06	52.88	1.50	5.50
					4.45	2.33	4.78	2.77	47.12	2.33	6.25
log	ham	fire	journal	homonym	4.96	2.79	4.68	4.04	56.28	2.17	5.00
					4.90	2.77	4.63	2.66	43.72	2.63	5.86
match	north	lighter	same	homonym	4.14	3.40	4.53	2.66	52.96	1.50	4.71
					4.14	3.51	2.64	4.33	47.04	1.20	6.25
mint	weep	coin	basil	homonym	4.54	2.44	4.89	2.70	51.28	3.13	7.00
					4.54	2.45	4.76	1.85	48.72	2.17	5.71
mold	lace	fungus	form	homonym	4.85	2.34	4.59	2.05	60.04	1.20	5.57
					4.85	2.28	3.13	3.34	39.96	3.33	5.50
mole	dorm	mouse	blemish	homonym	4.41	2.61	4.83	2.99	54.76	3.13	5.50
					4.41	2.57	4.07	1.49	45.24	2.67	5.14
mouth	radio	lip	opening	polyseme	4.74	3.73	4.96	2.74	57.84	1.17	6.13
					4.74	3.60	3.79	3.29	42.16	2.67	5.50
note	land	message	music	polyseme	4.61	3.44	3.97	3.67	50.84	1.33	5.43
					4.57	3.65	4.31	3.89	49.16	1.80	6.00
nut	cap	cashew	lunatic	homonym	4.52	2.90	4.92	1.08	63.04	1.17	5.25
					4.59	2.98	3.27	2.64	36.96	1.60	4.57

paper	hotel	cardboard	document	polyseme	4.93	3.72	4.90	2.14	56.72	2.33	6.38
					4.93	3.72	4.41	2.71	43.28	1.80	5.00
passage	crooked	journey	doorway	polyseme	3.80	2.59	2.57	3.01	56.80	2.50	5.25
					3.79	2.46	4.75	2.22	43.20	2.00	4.17
perch	groan	salmon	branch	homonym	4.10	1.71	4.81	2.53	57.48	4.25	6.88
					4.09	1.68	4.90	2.71	42.52	3.00	4.86
period	virgin	time	comma	homonym	3.31	3.15	3.07	5.00	54.00	3.00	5.50
					3.31	2.98	4.65	1.71	46.00	1.67	6.50
pillar	bleach	supporter	pedestal	polyseme	4.77	1.91	3.26	1.58	53.72	1.33	6.00
					4.74	2.08	4.48	1.83	46.28	2.80	6.33
pine	monk	flooring	sap	polyseme	4.37	2.50	4.43	1.36	50.08	2.67	5.29
					4.35	2.58	4.37	2.25	49.92	1.88	6.25
pipe	salt	plumbing	cigar	polyseme	4.88	3.00	4.24	2.39	51.88	2.60	6.25
					4.89	3.00	4.93	2.82	48.12	1.83	4.86
pitcher	diploma	jug	catcher	homonym	4.93	2.22	4.96	2.13	50.24	2.00	6.25
					4.93	2.11	4.44	2.28	49.76	1.63	5.67
pot	gym	marijuana	stove	homonym	4.81	3.06	4.89	2.43	61.04	1.20	5.43
					4.83	2.97	4.96	2.59	38.96	1.50	6.00
present	dressed	now	offering	homonym	3.39	3.66	1.48	5.21	50.40	1.00	5.00
					3.41	3.38	2.82	2.89	49.60	2.67	5.33
pupil	satin	eye	student	homonym	4.55	2.21	4.90	3.76	54.72	1.17	6.33
					4.57	2.13	4.92	3.34	45.28	1.20	6.63
racket	resort	shouting	paddle	polyseme	4.26	2.58	3.97	2.87	61.96	2.50	5.75
					4.30	2.55	4.80	2.28	38.04	2.33	3.00
root	hail	beginning	carrot	homonym	4.34	2.73	2.50	3.51	54.36	2.17	5.00
					4.32	2.79	5.00	2.29	45.64	2.25	6.25
sage	malt	wisdom	spice	homonym	4.54	1.95	1.53	2.75	56.92	3.80	6.88
					4.57	1.93	4.54	2.43	43.08	2.50	4.00
scale	stall	fish	pound	homonym	4.39	2.69	5.00	3.63	54.12	2.00	5.50
					4.37	2.66	4.61	2.85	45.88	2.17	6.38
sentence	ceremony	prison	paragraph	homonym	3.57	3.02	4.68	3.53	54.36	1.83	5.50
					3.57	2.91	3.96	2.16	45.64	1.80	4.17
sheet	badge	page	blanket	polyseme	4.93	2.77	4.90	3.28	56.24	2.00	4.33

					4.93	2.89	5.00	2.82	43.76	1.38	6.50
shower	turkey	storm	bath	polyseme	4.89	3.32	4.70	3.20	53.12	3.60	6.50
					4.89	3.06	4.85	3.20	46.88	1.00	5.67
sign	kids	clue	placard	polyseme	4.62	3.83	2.93	2.95	53.56	1.88	5.88
					4.61	4.19	4.46	0.85	46.44	2.20	6.00
space	folks	astronaut	territory	polyseme	3.54	3.53	4.75	2.31	52.88	1.25	5.38
					3.52	3.59	3.41	2.87	47.12	2.83	4.43
spell	awake	hex	interval	homonym	3.32	3.27	2.16	1.83	53.72	2.50	5.14
					3.32	3.13	2.57	1.66	46.28	4.50	6.75
spring	leader	summer	bounce	homonym	3.89	3.20	3.64	3.60	51.32	1.88	5.33
					3.89	3.20	3.86	2.70	48.68	2.67	6.25
staff	punch	teacher	stick	homonym	4.36	3.21	4.52	3.45	52.36	1.33	6.88
					4.39	3.18	4.59	3.70	47.64	2.20	4.83
straw	drums	tube	hay	polyseme	4.77	2.50	4.82	2.92	58.48	2.50	5.17
					4.79	2.57	4.87	2.51	41.52	2.60	6.25
stroke	potter	athlete	hospital	homonym	4.10	2.82	4.30	2.37	58.92	3.67	4.94
					4.12	3.02	4.64	3.80	41.08	1.88	6.56
stump	latch	tree	amputee	polyseme	4.78	2.10	5.00	3.52	51.96	1.63	6.13
					4.79	2.00	4.79	1.04	48.04	2.67	5.50
temple	jersey	church	forehead	homonym	4.53	2.95	4.90	3.55	57.04	2.00	6.00
					4.56	2.97	4.90	2.60	42.96	2.00	6.25
tin	bee	aluminum	container	polyseme	4.87	2.65	4.88	2.06	54.92	1.50	6.25
					4.88	2.72	4.85	2.34	45.08	2.20	4.29
tip	wet	waiter	point	homonym	4.50	3.15	4.67	2.83	51.76	1.67	4.57
					4.46	3.30	3.39	4.08	48.24	1.63	6.00
toast	cabin	croissant	tribute	homonym	4.93	3.23	4.96	1.66	50.12	2.00	6.63
					4.92	3.00	2.67	2.43	49.88	3.20	6.00
tongue	button	gum	dialect	polyseme	4.93	3.20	4.89	2.84	54.00	2.17	6.17
					4.96	3.16	2.67	1.84	46.00	1.88	6.25
trial	enter	attempt	court	polyseme	3.07	3.40	2.22	2.99	50.04	1.67	5.00
					3.12	3.18	4.31	3.71	49.96	1.40	3.71
trunk	couch	car	nose	homonym	4.71	3.00	4.89	4.39	53.16	1.33	6.13
					4.71	3.08	4.89	3.55	46.84	3.33	6.00

vessel	shield	nautical	bowl	polyseme	4.66	2.68	2.45	1.66	52.08	2.38	5.33
					4.66	2.62	4.87	3.04	47.92	3.83	5.50
volume	maniac	ear	amount	polyseme	3.07	2.55	5.00	3.21	54.04	1.60	5.83
					3.10	2.68	2.74	3.10	45.96	1.50	6.38

Table 33. Mean frequency and concreteness scores for Experiment 1A filler stimuli

Prime	Target	Target Type	Prime Freq.	Prime Conc.	Target Freq.	Target Conc.
roof	floor	related	3.26	4.79	3.71	4.8
mirror	image		3.09	4.97	3.06	3.85
juice	water		3.14	4.89	4.06	5
wing	feather		3.01	4.86	2.53	4.9
river	lake		3.45	4.89	3.26	4.88
road	way		3.76	4.75	4.86	2.34
face	mask		4.17	4.87	3.00	4.96
example	prototype		3.18	3.03	2.23	3.69
knight	nobleman		3.14	4.79	1.58	3.85
book	magazine		3.96	4.90	3.23	5
arrow	dart		2.60	4.97	2.00	4.9
cloud	sky		2.78	4.54	3.36	4.45
recovery	improvement		2.67	2.68	2.30	2.6
candle	torch		2.61	4.86	2.41	4.76
bone	fossil		3.12	4.90	1.90	4.9
color	red		3.30	4.08	3.88	4.24
monkey	gorilla		3.23	4.90	2.45	4.97
voice	speech		3.64	4.13	3.29	3.37
scar	stock	unrelated	2.64	4.74	3.11	4.19
spine	knitting		2.47	4.88	2.06	4.14
bird	vaccine		3.37	5.00	2.00	4.69
funeral	jingle		3.23	3.83	2.41	3.7
task	dare		2.81	2.84	3.45	2.28
meat	clown		3.35	4.90	2.91	4.9

tension	coffin		2.64	2.60	2.66	4.86
coward	epic		2.87	2.93	2.08	2.19
art	highway		3.56	4.17	2.96	4.72
trash	trick		3.06	4.70	3.38	3.36
ant	oval		2.44	4.86	1.85	4.04
fate	version		3.14	1.53	2.92	1.7
fog	clock		2.68	4.66	3.48	5
garlic	camel		2.49	4.89	2.41	4.93
mercy	role		3.11	1.57	2.97	2.19
pear	altar		1.84	4.93	2.42	4.85
stench	agent		2.06	3.85	3.72	3.61
truth	send		3.99	1.96	3.96	2.7
alley	veis	high-freq. nonword	2.92	4.82	1660.00	NA
	veir	low-freq. nonword	2.92	4.82	1010.00	NA
antenna	clar	high-freq. nonword	2.09	4.75	1854.33	NA
	clow	low-freq. nonword	2.09	4.75	911.33	NA
arms	orein	high-freq. nonword	3.48	4.97	3625.25	NA
	osoun	low-freq. nonword	3.48	4.97	1289.50	NA
ashes	rast	high-freq. nonword	2.70	4.92	2709.00	NA
	fack	low-freq. nonword	2.70	4.92	944.67	NA
axis	funed	high-freq. nonword	1.91	3.57	2256.25	NA
	dudge	low-freq. nonword	1.91	3.57	516.25	NA
blob	spong	high-freq. nonword	1.79	4.06	2607.25	NA
	spoff	low-freq. nonword	1.79	4.06	626.75	NA
bottom	rint	high-freq. nonword	3.41	4.25	4262.33	NA
	likh	low-freq. nonword	3.41	4.25	954.67	NA
bubbly	vit	high-freq. nonword	1.85	3.52	1477.00	NA
	ket	low-freq. nonword	1.85	3.52	1126.00	NA
bud	eatrit	high-freq. nonword	3.26	4.48	2547.00	NA
	outvot	low-freq. nonword	3.26	4.48	757.80	NA
bump	binar	high-freq. nonword	2.80	4.10	2975.75	NA
	bigyr	low-freq. nonword	2.80	4.10	407.25	NA
bush	steck	high-freq. nonword	2.86	4.90	2551.00	NA

	swuck	low-freq. nonword	2.86	4.90	396.50	NA
calculator	memenery	high-freq. nonword	1.83	4.86	2482.86	NA
	memixety	low-freq. nonword	1.83	4.86	902.14	NA
carbon	trore	high-freq. nonword	2.43	4.22	2781.75	NA
	shoff	low-freq. nonword	2.43	4.22	773.00	NA
clap	bleer	high-freq. nonword	2.38	4.16	2907.00	NA
	queed	low-freq. nonword	2.38	4.16	1575.00	NA
cleanup	bont	high-freq. nonword	2.00	3.04	2796.00	NA
	boft	low-freq. nonword	2.00	3.04	374.67	NA
clutch	dededer	high-freq. nonword	2.11	3.70	3822.50	NA
	demaxer	low-freq. nonword	2.11	3.70	1939.33	NA
comment	oal	high-freq. nonword	2.82	3.29	1742.00	NA
	oam	low-freq. nonword	2.82	3.29	682.50	NA
cones	cand	high-freq. nonword	1.79	4.57	2349.33	NA
	hawn	low-freq. nonword	1.79	4.57	512.00	NA
coupon	stip	high-freq. nonword	2.05	4.79	2956.00	NA
	twip	low-freq. nonword	2.05	4.79	368.67	NA
crow	foused	high-freq. nonword	2.36	4.93	2049.80	NA
	gourth	low-freq. nonword	2.36	4.93	1143.20	NA
delete	rop	high-freq. nonword	1.96	3.48	1447.00	NA
	pob	low-freq. nonword	1.96	3.48	717.00	NA
disarm	embere	high-freq. nonword	2.04	3.55	2600.60	NA
	empock	low-freq. nonword	2.04	3.55	911.00	NA
disaster	dreat	high-freq. nonword	2.95	3.07	2733.50	NA
	wheah	low-freq. nonword	2.95	3.07	966.00	NA
disk	flerm	high-freq. nonword	2.53	4.80	2658.50	NA
	spebe	low-freq. nonword	2.53	4.80	897.25	NA
dogs	scu	high-freq. nonword	3.43	5.00	768.00	NA
	swu	low-freq. nonword	3.43	5.00	101.00	NA
drake	cestar	high-freq. nonword	2.48	4.26	2905.60	NA
	cyllyr	low-freq. nonword	2.48	4.26	698.20	NA
dried	gectar	high-freq. nonword	2.41	3.54	1698.80	NA
	guxwar	low-freq. nonword	2.41	3.54	807.60	NA

drift	cont	high-freq. nonword	2.52	3.07	3365.67	NA
	coze	low-freq. nonword	2.52	3.07	1003.33	NA
duck	hain	high-freq. nonword	3.10	4.86	2986.33	NA
	rawl	low-freq. nonword	3.10	4.86	1140.33	NA
dungeon	bued	high-freq. nonword	2.12	4.32	1802.00	NA
	buft	low-freq. nonword	2.12	4.32	267.67	NA
eaten	weinter	high-freq. nonword	2.97	3.61	3544.00	NA
	wiegeer	low-freq. nonword	2.97	3.61	1733.00	NA
elf	sping	high-freq. nonword	2.29	4.30	3116.00	NA
	spiud	low-freq. nonword	2.29	4.30	515.50	NA
fall	vantet	high-freq. nonword	3.78	4.04	2545.40	NA
	zaffet	low-freq. nonword	3.78	4.04	545.60	NA
flash	yanst	high-freq. nonword	2.89	3.67	2254.50	NA
	yauze	low-freq. nonword	2.89	3.67	279.00	NA
gift	prip	high-freq. nonword	3.52	4.56	1581.33	NA
	swip	low-freq. nonword	3.52	4.56	396.67	NA
glide	ren	high-freq. nonword	2.08	3.93	4179.00	NA
	peb	low-freq. nonword	2.08	3.93	951.00	NA
glow	antind	high-freq. nonword	2.47	3.65	3987.00	NA
	attixt	low-freq. nonword	2.47	3.65	1960.20	NA
golfer	lert	high-freq. nonword	1.76	4.71	3527.00	NA
	lewn	low-freq. nonword	1.76	4.71	1296.67	NA
greasy	wallin	high-freq. nonword	2.30	3.82	2997.60	NA
	gyllod	low-freq. nonword	2.30	3.82	770.00	NA
grown	cangin	high-freq. nonword	3.11	3.20	3335.60	NA
	cunyod	low-freq. nonword	3.11	3.20	658.40	NA
gulf	anued	high-freq. nonword	2.38	4.08	2160.50	NA
	adynx	low-freq. nonword	2.38	4.08	327.00	NA
hernia	anain	high-freq. nonword	1.96	3.58	3181.00	NA
	apoke	low-freq. nonword	1.96	3.58	768.75	NA
hostess	inas	high-freq. nonword	2.28	4.12	3263.00	NA
	ipys	low-freq. nonword	2.28	4.12	286.67	NA
inn	prite	high-freq. nonword	2.63	4.64	2697.50	NA

	quiud	low-freq. nonword	2.63	4.64	347.50	NA
ivy	surfede	high-freq. nonword	2.24	4.50	1667.50	NA
	surfoke	low-freq. nonword	2.24	4.50	677.50	NA
king	cug	high-freq. nonword	3.82	4.10	487.00	NA
	bup	low-freq. nonword	3.82	4.10	461.50	NA
lark	meding	high-freq. nonword	1.92	4.28	3783.40	NA
	sozing	low-freq. nonword	1.92	4.28	2350.80	NA
lined	cery	high-freq. nonword	2.52	3.59	2809.33	NA
	cyxy	low-freq. nonword	2.52	3.59	65.67	NA
lodge	ain	high-freq. nonword	2.53	4.00	3907.00	NA
	akh	low-freq. nonword	2.53	4.00	178.50	NA
mall	setrince	high-freq. nonword	2.98	4.83	2553.14	NA
	seglance	low-freq. nonword	2.98	4.83	1633.29	NA
melt	tuli	high-freq. nonword	2.57	3.96	1502.33	NA
	huxi	low-freq. nonword	2.57	3.96	149.00	NA
merchant	leam	high-freq. nonword	2.17	4.31	2058.67	NA
	feak	low-freq. nonword	2.17	4.31	905.67	NA
mesa	santer	high-freq. nonword	1.83	3.77	3558.80	NA
	ruoyer	low-freq. nonword	1.83	3.77	1444.60	NA
moist	minn	high-freq. nonword	1.97	4.00	2877.67	NA
	mulk	low-freq. nonword	1.97	4.00	499.00	NA
mush	preat	high-freq. nonword	2.16	3.67	2970.50	NA
	swoot	low-freq. nonword	2.16	3.67	591.25	NA
nation	rart	high-freq. nonword	3.02	3.62	2359.33	NA
	roke	low-freq. nonword	3.02	3.62	1090.00	NA
nine	ainstay	high-freq. nonword	3.54	4.04	2589.17	NA
	alfskay	low-freq. nonword	3.54	4.04	660.67	NA
ninth	derd	high-freq. nonword	2.53	3.57	3130.33	NA
	dosk	low-freq. nonword	2.53	3.57	553.00	NA
notch	rearse	high-freq. nonword	2.23	4.23	2564.40	NA
	heague	low-freq. nonword	2.23	4.23	1025.60	NA
pansy	shate	high-freq. nonword	2.00	3.89	2754.00	NA
	shaye	low-freq. nonword	2.00	3.89	761.25	NA

peddle	rallet	high-freq. nonword	1.74	4.04	2512.20	NA
	dullot	low-freq. nonword	1.74	4.04	1077.00	NA
pest	kel	high-freq. nonword	2.17	3.96	1316.00	NA
	kek	low-freq. nonword	2.17	3.96	481.00	NA
pew	mon	high-freq. nonword	1.73	3.92	2785.00	NA
	wob	low-freq. nonword	1.73	3.92	348.50	NA
piano	certy	high-freq. nonword	3.10	4.90	2374.75	NA
	cekry	low-freq. nonword	3.10	4.90	589.00	NA
plug	bant	high-freq. nonword	2.73	4.64	2426.00	NA
	bazz	low-freq. nonword	2.73	4.64	318.33	NA
pull	ral	high-freq. nonword	3.87	3.97	3087.00	NA
	paj	low-freq. nonword	3.87	3.97	558.00	NA
putty	engring	high-freq. nonword	1.97	4.48	3695.17	NA
	ewsbing	low-freq. nonword	1.97	4.48	1983.83	NA
race	conx	high-freq. nonword	3.50	3.59	2363.67	NA
	cown	low-freq. nonword	3.50	3.59	1108.00	NA
rebel	beant	high-freq. nonword	2.44	3.07	2293.50	NA
	beamt	low-freq. nonword	2.44	3.07	905.75	NA
reflex	bule	high-freq. nonword	2.07	3.10	1656.33	NA
	buke	low-freq. nonword	2.07	3.10	480.00	NA
removal	drem	high-freq. nonword	2.03	3.00	1983.00	NA
	snum	low-freq. nonword	2.03	3.00	356.00	NA
riding	cedcus	high-freq. nonword	3.21	4.14	1655.00	NA
	cymcus	low-freq. nonword	3.21	4.14	490.40	NA
rip	ralt	high-freq. nonword	3.01	3.79	2181.67	NA
	saft	low-freq. nonword	3.01	3.79	376.00	NA
roster	bintom	high-freq. nonword	2.05	3.76	2622.40	NA
	birtup	low-freq. nonword	2.05	3.76	714.40	NA
row	rurer	high-freq. nonword	3.13	3.93	3180.50	NA
	muyer	low-freq. nonword	3.13	3.93	1678.25	NA
rub	beath	high-freq. nonword	2.90	4.33	2044.25	NA
	booge	low-freq. nonword	2.90	4.33	851.75	NA
scan	fleer	high-freq. nonword	2.68	3.48	2730.25	NA

	flooh	low-freq. nonword	2.68	3.48	728.75	NA
scary	pung	high-freq. nonword	3.13	3.00	2057.00	NA
	lult	low-freq. nonword	3.13	3.00	662.33	NA
scope	wint	high-freq. nonword	2.46	3.74	3460.33	NA
	wiud	low-freq. nonword	2.46	3.74	293.33	NA
sea	rit	high-freq. nonword	3.48	4.79	2525.50	NA
	rab	low-freq. nonword	3.48	4.79	2058.00	NA
sell	lierer	high-freq. nonword	3.67	3.35	4173.40	NA
	loaker	low-freq. nonword	3.67	3.35	1856.00	NA
set	yest	high-freq. nonword	4.07	3.41	2745.33	NA
	veke	low-freq. nonword	4.07	3.41	957.67	NA
sever	treel	high-freq. nonword	1.91	4.00	2208.50	NA
	sweel	low-freq. nonword	1.91	4.00	821.00	NA
shade	bringe	high-freq. nonword	2.48	3.38	3094.80	NA
	brudge	low-freq. nonword	2.48	3.38	589.00	NA
shell	meer	high-freq. nonword	2.83	4.80	2924.33	NA
	deeg	low-freq. nonword	2.83	4.80	1301.67	NA
shiny	smersed	high-freq. nonword	2.60	3.33	2735.67	NA
	smelfth	low-freq. nonword	2.60	3.33	930.83	NA
skating	latle	high-freq. nonword	2.40	4.75	2509.00	NA
	hawhe	low-freq. nonword	2.40	4.75	838.00	NA
sketch	pon	high-freq. nonword	2.40	4.56	2840.50	NA
	pav	low-freq. nonword	2.40	4.56	735.50	NA
skinny	manter	high-freq. nonword	2.85	3.57	3747.60	NA
	munner	low-freq. nonword	2.85	3.57	2203.00	NA
sling	eraden	high-freq. nonword	2.07	4.52	3348.00	NA
	ecaxen	low-freq. nonword	2.07	4.52	1420.60	NA
sofa	tond	high-freq. nonword	2.48	4.90	2578.00	NA
	tolt	low-freq. nonword	2.48	4.90	1011.67	NA
spook	seriom	high-freq. nonword	2.15	3.14	2958.40	NA
	sefiug	low-freq. nonword	2.15	3.14	730.00	NA
squat	rem	high-freq. nonword	2.21	4.32	2776.50	NA
	wim	low-freq. nonword	2.21	4.32	677.00	NA

stag	reron	high-freq. nonword	1.91	4.39	4363.25	NA
	meyon	low-freq. nonword	1.91	4.39	1667.75	NA
steel	liel	high-freq. nonword	2.97	4.87	1911.67	NA
	guez	low-freq. nonword	2.97	4.87	273.33	NA
stood	brate	high-freq. nonword	3.12	3.41	3065.25	NA
	slaff	low-freq. nonword	3.12	3.41	771.00	NA
stormy	bist	high-freq. nonword	1.88	3.96	2313.33	NA
	bift	low-freq. nonword	1.88	3.96	426.33	NA
stride	bleset	high-freq. nonword	2.03	3.81	2494.00	NA
	swopet	low-freq. nonword	2.03	3.81	860.40	NA
swipe	tedlet	high-freq. nonword	1.92	4.26	2792.40	NA
	tuxlet	low-freq. nonword	1.92	4.26	1108.20	NA
tab	minth	high-freq. nonword	2.47	4.14	3152.00	NA
	moque	low-freq. nonword	2.47	4.14	481.50	NA
teller	grorm	high-freq. nonword	2.12	4.38	1557.00	NA
	spowl	low-freq. nonword	2.12	4.38	668.75	NA
toll	pange	high-freq. nonword	2.24	3.54	2427.75	NA
	pairn	low-freq. nonword	2.24	3.54	824.75	NA
tomato	coath	high-freq. nonword	2.48	5.00	2099.25	NA
	woach	low-freq. nonword	2.48	5.00	934.75	NA
town	reringtent	high-freq. nonword	4.10	4.64	3961.00	NA
	revempment	low-freq. nonword	4.10	4.64	1934.67	NA
tub	eiste	high-freq. nonword	2.81	4.64	2757.25	NA
	euque	low-freq. nonword	2.81	4.64	274.50	NA
tumor	pess	high-freq. nonword	2.42	4.56	2574.33	NA
	pift	low-freq. nonword	2.42	4.56	494.67	NA
walrus	runting	high-freq. nonword	1.76	5.00	3521.50	NA
	mudging	low-freq. nonword	1.76	5.00	2040.50	NA
watt	stang	high-freq. nonword	1.99	3.10	3211.00	NA
	stawn	low-freq. nonword	1.99	3.10	1458.25	NA

B.2 EXPERIMENT 1B

Table 34. Mean familiarity, figurative dominance, meaning relatedness, and target frequency/concreteness scores from Experiment 1B idiom stimuli

Idiom Prime	Control Prime	Literal Target	Figurative Target	Idiom Fam.	Figurative Dominance	Meaning Relatedness	Lit. Target Freq.	Lit. Target Conc.	Fig. Target Freq.	Fig. Target Conc.
add fuel to the fire	take tools to the garage	camping	worsen	1.15	82.75	3.25	1.32	1.70	2.50	4.00
asleep at the wheel	stuck in the elevator	driver	slacking	2.15	62.86	3.64	1.60	2.36	3.38	4.71
at the end of your rope	in the pocket of your bag	lasso	frustrate	2.85	87.36	3.43	1.18	1.82	1.66	4.74
bent out of shape	moved out of town	distorted	frazzle	1.45	57.86	3.25	0.85	2.20	2.02	2.57
blow a fuse	book a flight	electricity	fury	1.46	63.86	4.36	2.29	2.89	2.77	3.90
blow off some steam	look at those trees	hydraulics	anger	1.09	84.00	3.92	3.00	2.41	1.46	3.71
break a leg	throw a ball	doctor	lucky	1.38	83.83	3.64	3.86	1.76	4.13	4.69
break your back	irritate your sibling	paralysis	labor	1.85	66.25	5.08	2.79	3.08	2.00	3.52
bring home the bacon	step onto the platform	groceries	breadwinner	2.23	76.64	5.00	1.23	3.24	2.48	4.74
change your tune	clean your window	whistle	reverse	2.77	66.86	4.29	2.75	3.38	2.90	4.42
cover your tracks	raise your hands	footprint	disguise	1.54	77.79	3.36	2.60	3.97	1.75	4.37
crack the whip	fill the pie	horse	authority	2.15	77.71	2.75	3.03	2.34	3.68	5.00
cut down to	stay up to	trim	belittle	2.54	61.67	4.92	1.58	2.10	2.34	3.93

size	date									
deliver the goods	water the garden	mailbox	fulfill	2.83	59.14	2.67	2.30	1.78	2.33	5.00
fall off the wagon	sit on the chair	tumble	relapse	2.54	78.33	4.29	1.68	2.21	1.90	3.89
go like clockwork	smell like cheese	mechanics	efficient	2.92	82.25	2.71	2.40	1.82	2.09	3.59
go with the flow	take down the curtains	downstream	relax	1.08	85.14	2.17	3.69	2.86	1.81	4.19
have a heart	sharpen a pencil	artery	charity	1.77	81.58	3.08	2.93	2.62	2.43	4.48
have cold feet	need bright light	shiver	reluctant	1.62	66.93	3.21	1.90	3.88	2.11	1.76
hit the bullseye	buy the milk	arrow	exact	1.38	65.17	1.92	3.06	2.54	2.60	4.97
hit the nail on the head	throw the ball over the fence	construction	precision	1.15	85.21	2.43	2.18	2.41	2.85	3.72
hit the sack	touch the grass	baggage	snore	1.08	89.64	4.50	1.92	4.39	2.50	4.43
hold your horses	feed your children	reins	patience	1.67	79.93	2.58	2.89	1.66	1.88	4.56
in the hole	on the wall	bury	debt	2.55	78.79	2.92	2.86	2.72	3.02	3.82
jump the gun	take the purse	sprint	premature	1.31	72.42	2.67	2.25	2.72	1.57	4.14
jump through hoops	bury under sand	acrobat	requirement	1.69	80.36	2.50	1.87	2.52	1.49	4.46
leave no stone unturned	let some people inside	pebble	thorough	2.55	84.07	2.42	2.38	1.66	1.82	4.86
let it rip	turn it over	shred	release	1.85	84.21	4.67	3.27	3.24	2.18	4.38
let your hair	put your	ponytail	freedom	1.70	69.29	3.83	3.23	2.34	1.72	4.77

down	jacket on									
live on the edge	shop in the mall	cliff	reckless	1.46	82.75	2.93	2.40	2.36	3.04	4.71
lose your grip	open your wallet	gymnast	lapse	1.69	63.08	3.17	1.85	2.85	1.49	4.85
make the cut	write a poem	incision	audition	2.54	71.08	4.36	2.85	3.66	2.09	4.07
miss the mark	pick the color	archery	blunder	1.62	68.33	2.25	1.67	2.35	1.61	3.93
on the fence	by the river	railing	decision	1.18	89.00	2.25	3.45	2.19	1.88	4.79
out of gas	out of sight	refuel	tired	2.54	50.17	3.07	3.76	3.00	1.64	3.53
over the hill	in the line	hike	older	2.23	56.67	2.92	3.33	3.19	2.52	4.14
play by ear	work with vigor	imitate	improvise	1.69	61.25	4.07	2.11	2.07	1.97	2.66
play with fire	talk to her	ignite	hazard	1.36	72.64	2.17	2.30	3.67	1.72	4.04
pull the plug	wash the puppy	power	stop	1.23	76.14	4.14	4.56	3.68	3.88	2.04
put your cards on the table	leave your laundry in the machine	gamble	honest	1.92	70.86	2.93	3.57	1.66	2.66	3.17
put your foot down	touch my makeup up	stomp	stubborn	1.85	71.25	3.86	2.74	2.18	2.14	4.41
raise the roof	fill the pail	blueprint	party	1.69	84.67	3.50	4.08	3.89	1.81	4.77
ring a bell	eat a sandwich	music	remember	1.23	69.07	3.43	4.44	2.41	3.89	4.31
rock the boat	pet the dog	wave	disturb	1.92	77.75	3.71	2.75	3.04	3.04	4.55
run out of steam	run through the woods	machine	drained	1.36	88.79	1.75	2.05	3.23	3.55	4.25
see the light	direct a play	illuminate	realize	1.18	73.57	2.25	3.61	2.03	1.63	3.55
shut your	move your	snare	voice	1.18	81.14	2.75	3.64	4.13	1.60	3.90

trap	car									
sing the blues	walk the dogs	microphone	depression	2.54	57.67	4.25	2.61	2.39	2.37	4.88
skate on thin ice	speak with great passion	freeze	danger	1.50	78.86	2.50	3.35	2.68	3.22	3.96
slap in the face	place on the table	palm	insult	1.46	66.79	4.79	2.81	2.79	2.83	4.83
spill the beans	drop the hat	mess	tattle	1.31	85.08	3.25	0.95	2.69	3.60	3.90
stick your nose into	hand your cash over	sniff	meddle	1.77	78.29	3.43	1.81	2.43	2.26	4.17
take for a ride	put down the box	passenger	cheat	2.69	54.58	4.75	2.74	4.34	2.96	2.23
take the cake	remove the chair	bakery	winner	2.15	79.50	5.86	3.20	3.21	2.26	4.83
take with a grain of salt	walk with a bottle in hand	flavor	caution	1.85	87.71	3.71	2.42	2.04	2.41	3.55
tear your hair out	take your shoes off	tweezers	tantrum	2.38	81.93	3.00	1.66	3.37	1.72	4.96
throw in the towel	copy onto the paper	laundry	defeat	1.46	81.17	2.00	2.76	2.96	2.99	4.93
turn back the clock	take out the screw	reset	history	2.46	61.17	3.71	3.63	2.96	2.22	3.00
turn over a new leaf	run across the new bridge	gardener	rehab	2.09	86.50	2.50	2.52	3.39	2.33	4.50
turn your back	wash your face	pivot	abandon	1.77	74.14	3.14	2.62	2.54	1.38	3.72

Table 35. Mean target frequency and concreteness scores for Experiment 1B filler stimuli

Prime	Related/High Freq. Target	Unrelated/Low Freq. Target	Related/High - Frequency	Related/High - Concreteness	Unrelated/Low - Frequency	Unrelated/Low - Concreteness
about a month ago	time	take	5.00	3.07	4.98	3.06
around the mountain	path	gang	3.10	4.41	3.19	4.43
at the beginning	start	looks	4.24	2.71	4.20	2.70
back in the car	drive	class	3.89	3.86	3.78	3.85
bought the car	dealer	stable	2.92	3.76	2.83	3.77
cost an arm and a leg	expensive	backwards	3.15	3.13	2.82	3.17
don't care about	lazy	obey	2.77	2.67	2.66	2.67
drive to the bank	money	night	4.51	4.54	4.65	4.52
far from the city	suburb	lather	1.49	3.76	1.46	3.77
fill up the bottle	drink	child	4.10	4.76	3.91	4.78
finish the project fast	complete	selected	3.42	2.70	2.40	2.68
flap its wings	feather	mansion	2.53	4.90	2.52	4.89
get it together	organize	deserted	2.34	2.72	2.44	2.72
head in the clouds	distract	extended	2.40	2.33	2.33	2.34
hit the road	journey	confess	3.01	2.57	2.91	2.57
hit the spot	perfect	feeling	3.91	1.69	3.93	1.68
in a tight spot	stuck	wrote	3.53	3.55	3.56	3.55
in your spare time	leisure	remorse	2.11	2.03	2.19	2.00
know how long	clock	smoke	3.48	5.00	3.52	4.96
learn an instrument	orchestra	submarine	2.45	4.79	2.56	4.80
light a candle	flame	smack	2.66	4.67	2.69	4.67
lose your temper	furious	amateur	2.49	2.31	2.53	2.31
not a bad person	kind	stay	4.48	2.07	4.42	2.15
open the door	hinge	gnome	1.48	4.57	1.52	4.59
out of the way	clear	state	3.94	3.55	3.74	3.52
out of your league	unprepared	continuous	1.90	2.07	2.00	2.11
out on the lake	rowboat	machete	1.74	4.81	1.73	4.82
shed some light	explain	mistake	3.75	1.97	3.72	1.97
sit in your chair	desk	beer	3.35	4.87	3.59	4.88
swallow your pride	humble	reckon	2.70	1.73	2.78	1.74

take care of things	errand	linked	2.31	3.37	2.34	3.36
tip of the iceberg	uncover	babysit	2.00	3.42	2.02	3.45
turn the tables	flip	pink	2.86	3.97	3.16	3.93
watch the stars	telescope	pepperoni	2.18	5.00	2.12	5.00
when we go home	garage	mirror	3.14	4.96	3.09	4.97
wrapped around your finger	manipulate	functional	2.25	2.55	2.15	2.55
a cup of tea	selicin	pivicid	2781.00	NA	1097.50	NA
a drink of milk	grenting	bludging	3959.29	NA	1949.71	NA
a fashionable dress	sustled	metbred	2369.00	NA	2126.33	NA
a lecture about music	rewinsta	rekarque	2787.43	NA	1233.86	NA
a lot of juice	wondy	gippy	1714.00	NA	423.75	NA
a lot of noise	enstasiasm	erphusiasm	1929.11	NA	1504.00	NA
a piece of cheese	perente	sedanso	3932.67	NA	2195.17	NA
a restricted diet	ponten	raffen	3382.00	NA	1626.40	NA
above the trees	corractible	corhyptible	2128.50	NA	1579.80	NA
across the street	nond	noke	2389.00	NA	642.00	NA
adopt a cat	cathestal	cawreblol	2756.38	NA	1359.50	NA
and now the rest	tonute	tunups	2316.00	NA	715.40	NA
at that market	miseage	miquage	1662.00	NA	729.83	NA
back in the box	anant	afish	2852.75	NA	1271.50	NA
back to school	imecine	imeyize	2350.00	NA	687.00	NA
back to the grind	wrown	knohm	778.25	NA	257.50	NA
beat a dead horse	bedmer	bedmox	2665.60	NA	1297.00	NA
beginning to smell	bustervep	butterfaw	2362.50	NA	1684.63	NA
bite the bullet	remoin	remoft	2752.40	NA	1375.40	NA
bite the dust	dinch	wized	2921.75	NA	1513.25	NA
bridge the gap	prite	whibs	2697.50	NA	484.00	NA
buried in it all	listerran	ficherran	3350.00	NA	2461.38	NA
cash in your chips	hesple	hucque	2252.00	NA	340.80	NA
choose your battles	estel	etlyl	3545.75	NA	853.00	NA
close to the sea	cound	roobs	1941.50	NA	913.00	NA
cloud in the sky	mererity	mepucity	3480.00	NA	1017.14	NA

dead on your feet	morted	muoyed	2720.20	NA	1047.60	NA
don't get along	drach	smimp	1621.75	NA	915.50	NA
don't really need	mistain	suntaws	2897.83	NA	1324.83	NA
dressed to kill	prote	swove	2196.75	NA	744.00	NA
drop the anchor	dist	jodg	2733.67	NA	272.33	NA
eat your words	decoules	debearns	2427.29	NA	1543.86	NA
fail the test	invinint	ingitunt	3747.71	NA	2742.43	NA
find the key	destery	sovvery	3641.00	NA	1670.50	NA
first out of the gate	terricle	himbible	2899.43	NA	1144.00	NA
flying home	frostle	swaggle	1794.50	NA	955.83	NA
frog in your throat	tetel	leket	2969.50	NA	1433.00	NA
full of people	coose	loofs	1568.75	NA	657.75	NA
get a divorce	restiet	requeet	3330.00	NA	1304.33	NA
get your wires crossed	enserselent	ensaftsment	2934.90	NA	1690.70	NA
go back to school	bererious	bemucious	3281.13	NA	1163.25	NA
go for a walk	martist	pambost	2791.50	NA	1274.50	NA
go to the top	nart	naft	1798.33	NA	606.33	NA
grind to a halt	anysede	anysype	2165.00	NA	979.50	NA
hand in the till	rerast	resyls	3766.20	NA	1953.40	NA
have a cow	proup	knoup	1391.75	NA	746.00	NA
have a dog in the fight	strenic	phlonib	2903.50	NA	1435.00	NA
having kittens	visting	vodging	3782.00	NA	2055.17	NA
head in the sand	remp	tekg	2143.67	NA	1487.67	NA
heal the wound	borst	bugue	2138.00	NA	403.75	NA
hope you're doing well	jisor	jibyr	1584.75	NA	142.50	NA
hurt in an accident	narary	nafafy	2220.20	NA	486.60	NA
in a pickle	peretion	bevotood	3658.86	NA	760.86	NA
in front of the house	refrint	redrush	3025.67	NA	2141.67	NA
in the race	correrb	corrynx	2795.17	NA	1091.00	NA
in the weeds	sattery	sybvery	2796.67	NA	1498.33	NA

keep up your end	sesitine	mipitive	3477.43	NA	1778.14	NA
know what to do	proress	blofess	2807.17	NA	1580.00	NA
know what you mean	cooten	buiden	2503.00	NA	1643.40	NA
lift a finger	hean	hoer	2349.00	NA	2500.67	NA
load up the pallets	coute	teels	2371.25	NA	1853.50	NA
lose all the tools	dister	mynner	3751.60	NA	1799.60	NA
make some dinner	prare	clish	2990.25	NA	1845.25	NA
meet your eye	conretion	coywetion	3119.13	NA	2047.25	NA
miss by a mile	commerted	cormythed	2732.25	NA	1689.25	NA
move the sofa around	rerastion	rece	3840.25	NA	2492.00	NA
muddy the waters	jarement	joftment	2459.29	NA	1310.14	NA
no one knew	harer	wojer	3703.50	NA	1666.25	NA
not going to see	callin	bottad	3225.40	NA	1053.80	NA
out of date	derudion	pebygion	2682.14	NA	1383.71	NA
paint the town red	froticine	clivicive	2581.38	NA	1377.38	NA
piled on the floor	inatic	ivonic	3917.40	NA	1912.00	NA
plant in the ground	porein	sovoun	3115.20	NA	1022.00	NA
play the piano	sestery	mequecy	3546.00	NA	713.17	NA
predict the weather	diviste	divaque	2372.67	NA	696.67	NA
prepare for the earthquake	arrert	arrypt	3050.00	NA	930.80	NA
prescribe some pills	hutin	tyban	3226.00	NA	1252.00	NA
pull the reins	glan	smip	1979.33	NA	808.00	NA
pull up stakes	midteak	midgook	1462.50	NA	645.33	NA
raise an eyebrow	broravy	czamapy	1495.67	NA	626.17	NA
raised by wolves	piretulous	pimavulous	1545.78	NA	1048.00	NA
sail close to the wind	feated	hiewed	3089.40	NA	1523.40	NA
scaring the puppy	parmert	parmews	2258.67	NA	1134.67	NA
see the doctor	precert	prebuke	2664.17	NA	1260.83	NA
sell cheaper vegetables	eiss	ooch	1546.67	NA	1030.67	NA

show your hand	resern	rebext	3539.40	NA	1273.40	NA
shy and quiet	riste	rique	3398.25	NA	963.00	NA
sit in your truck	phink	gnunk	2204.75	NA	664.25	NA
slice the melon	atatier	avaloor	3738.50	NA	1510.50	NA
smash the bug	mawisem	sukimum	1362.50	NA	530.33	NA
stir the pot	deelint	deozist	2977.67	NA	1554.17	NA
take the bull by the horns	culle	cugue	1681.25	NA	438.50	NA
takes a lot	prorintion	profowtion	3435.44	NA	1874.11	NA
tell you a story	tandein	fashoun	2821.33	NA	1316.67	NA
the last moment	tont	cuke	2996.67	NA	526.33	NA
through the wringer	raren	rayen	3608.25	NA	1882.75	NA
throughout the land	unaintly	udaquely	2296.71	NA	732.86	NA
time for a nap	surfert	salvike	1680.67	NA	943.67	NA
try to help everyone	momple	rimque	1450.40	NA	941.20	NA
trying to show	sursestent	papsequent	2777.44	NA	1355.78	NA
under your feet	rond	voft	2837.00	NA	208.67	NA
up in the air	unisersal	upibizzal	2449.13	NA	767.00	NA
upset the apple cart	itrersely	imbiquely	2670.25	NA	775.38	NA
very well trained	barery	bampby	3016.00	NA	537.80	NA
water in the lake	brerorio	czemorio	3079.71	NA	1517.29	NA
when you were young	barelity	napolity	2235.00	NA	1471.43	NA
write a long letter	dencatery	denmoxery	3123.75	NA	1806.00	NA
youth and adults	shate	shaze	2754.00	NA	762.00	NA

B.3 EXPERIMENT 2

Table 36. Experiment 2 sentence stimuli and mean literal/figurative context bias scores

Literal Context / Figurative Context	Precritical Region	Critical Idiom	Postcritical Region	Wrap-Up Region	Literal Bias Score / Figurative Bias Score
Carol tossed more logs into the stove. Carol made the argument worse.	She always	adds fuel to the fire	when it isn't necessary.	We don't know why.	0.50 1.36
James drove home after a long night at the office. James could not focus on the board meeting.	He was	asleep at the wheel	for a little while.	He is okay now.	0.73 1.33
Eric is rock-climbing in Yosemite. Eric can't keep up with his nephew anymore.	He is	at the end of his rope	and unsure what to do.	He is exhausted.	1.83 1.27
Jack plugged in too many appliances. Jack is extremely angry.	He might	blow a fuse	if he's not careful.	Nothing happened.	0.73 1.25
Oliver opened up the pressure cooker. Oliver couldn't help punching the couch.	He needed to	blow off some steam	before he left.	It didn't help much.	1.25 1.27
Alex used the same theme song for several years. Alex used to be strongly anti-climate change.	Finally she	changed her tune	after lots of thought.	No one expected it.	2.36 1.00
Janice opened up the windows. Janice started discussing what happened.	She wanted to	clear the air	in the room.	It helped.	0.58 1.36

Sam placed leaves over the trail.	He had to	cover his tracks	before he left.	Nobody noticed it.	2.00
Sam deleted the files on his computer.					1.92
Amy works in a rodeo show.	When she	cracks the whip	everyone pays attention.	She's good at her job.	0.83
Amy is a controlling boss.					1.36
Bill tailors his pants to fit better.	He decided he will	cut them down to size	tomorrow morning.	He's an interesting person.	0.55
Bill was frustrated with his coworkers.					1.25
David left the groceries on the front porch.	He has to	deliver the goods	or he'll be fired.	People are waiting.	1.50
David has not played well in the games so far.					1.82
Luke helped out at a hay ride.	Last night he	fell off the wagon	and hurt himself.	It was scary.	0.18
Luke was struggling to stay sober.					1.33
Sue rowed her raft down the river.	She chooses to	go with the flow	as much as possible.	It is simple to do.	1.08
Sue agreed with her coworkers.					1.64
Ben wears thick socks all the time.	He always	gets cold feet	no matter what.	It's very frustrating.	0.42
Ben cancels our plans all the time.					1.45
Avery had very good aim.	She skillfully	hit the bulleseye	during her turn.	It was impressive.	0.45
Avery made a perfect guess.					1.58
Taylor was repairing the roof.	She	hit the nail on the head	with perfect precision.	It was very skillful.	0.83
Taylor answered a tough question during her dissertation defense.					1.45
Samir took his livestock to a veterinarian.	He had to	hold his horses	so the doctor could do her job	Everything was fine.	1.27
Samir wanted to get out of the hospital as soon as he could.					1.08
Fred's bosses at the circus give him a lot of work.	They always make him	jump through hoops	because they can.	It's ridiculous.	3.42
Frank will do anything to impress women.					1.27

Danny had almost worn through the knee of his jeans. Danny had to belch in the middle of dinner.	He decided to	let it rip	And not worry.	It all worked out.	0.45 1.08
Katie's ponytail was very uncomfortable. Katie is very anxious in social situations.	She decided to	let her hair down	and have fun.	She felt much better.	0.58 2.82
Peter is an expert on the parallel bars. Peter often panics about small things.	Sometimes he	loses his grip	for a moment.	But he always recovers.	0.91 1.83
Emily loved to use her scissors. Emily was the best applicant for the job.	When she	made the cut	no one was surprised.	She's very skilled.	0.50 1.45
It was the first round of the archery tournament. All the students had to take a pop quiz.	Sacha worked hard, but	missed the mark	on every try.	It was a bad day.	3.27 1.58
Alan's cat gets stuck in high places all the time. Alanna couldn't decide between two great job offers.	Yesterday she was	on the fence	for several hours.	It was tough.	0.33 1.45
Gabby's car stopped in the middle of the freeway. Gabby had been working hard all day.	She was	out of gas	and wanted to go home.	It was a bad situation.	0.27 2.17
Andrea was looking forward to the view at the end of her hike. Andrea couldn't wait to grow up.	But when she was	over the hill	she was disappointed.	It wasn't what she expected.	0.33 1.91
Jordan just heard a new song for the first time. Jordan doesn't practice his	But he can	play it by ear	quickly and easily.	He is very skilled.	0.55 1.33

speeches before he gives them.					
Mike is a professional juggler. Mike is illegally importing goods	He loves to	play with fire	and enjoys his work.	It's very dangerous.	2.50 1.36
Matt's circuit breaker is almost overloaded. Matt's research project is costing money and finding no results.	He should	pull the plug	before it's too late	It's a safe decision.	1.45 1.08
Omar got his foot stuck under a tree root. Omar is a very calm person.	Mel tried to	pull his leg	but he didn't appreciate it.	It didn't work.	0.58 2.18
Azia is a famous stage magician. Azia is negotiating an important contract.	She always	puts her cards on the table	so everyone can see.	She's very talented.	2.36 1.25
Joshua has a tendency to get sea sick. Joshua is nervous about pitching his research proposal.	He does not want anybody to	rock the boat	tomorrow morning.	Everybody understands.	0.83 1.55
Alice hadn't caught any animals. Alice's boss reprimanded her for talking too much.	She had to	shut her trap	and go home.	It was frustrating.	1.18 1.17
Jason is a talented singer. Jason is always complaining.	Every evening he	sings the blues	for us all.	It lasts a long time.	0.42 2.09
Kasey doesn't care that the pond isn't fully frozen. Kasey is mediating a heated argument between her parents.	We hold our breath as she	skates on thin ice	very recklessly.	We almost can't watch!	1.09 1.00
Dan thought that his girlfriend was weak. Dan thought that his boss would praise his work.	He was shocked by the	slap in the face	that she gave him.	He deserved it.	1.33 1.82
Mackenzy was making her favorite tacos. Mackenzy was working on a	Unfortunately, she	spilled the beans	before she finished.	What a disappointment.	0.82 1.33

surprise for her friend.					
Nikki's cat likes to sniff things. Nikki loves to gossip.	Sometimes it's frustrating that she	sticks her nose into	everything she sees.	But it doesn't matter much.	0.83 1.45
Jack just bought a new car. Jack is such a prankster.	He loves to	take his friends for a ride	whenever they get together.	He looks forward to seeing them.	0.55 2.00
Shelby just got home from the salon. Shelby is really stressed about her classes.	She wants to	tear her hair out	and throw things.	She'll get used to it.	3.58 2.64
Megan was washing her sweaters. Megan was jealous of her friends who were relaxing.	She decided to	throw in the towel	along with them.	It was that kind of day.	1.45 2.17
Marissa is scared of the haunted house. Marissa is involved with several progressive causes.	She decided not to	turn her back	on anything she saw.	She felt better.	1.45 1.33
Consuela had Michael in a headlock. Michael didn't agree with Consuela's new proposal.	Consuela	twisted his arm	until he gave in.	She was triumphant.	0.58 2.09
Colleen just took her first hot air balloon ride. Colleen had to decide between two great colleges.	She was	up in the air	for a long time.	It was very stressful.	1.36 1.00
Marcus just got a new dog. Marcus always teases his little sister.	He thinks it's funny to	yank her chain	to annoy her.	It's really not funny.	1.83 1.45

BIBLIOGRAPHY

Ammer, C. (Ed.) (2013) The American Heritage Dictionary of Idioms. Houghton Mifflin Harcourt.

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, 1-27. doi: 10.1080/23273798.2016.1171366

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *J Mem Lang, 62*(1), 67-82. doi: 10.1016/j.jml.2009.09.005

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

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J Mem Lang, 68*(3), 255-278. doi: 10.1016/j.jml.2012.11.001

Bates, DM, Maechler, M, Bolker, B, & Walker, S. (2016). lme4: Mixed-effects modeling with R; 2010. *URL:* http://lme4*. r-forge. r-project. org/book [8 April 2015]*.

Brocher, A., Koenig, J-P., & Foraker, S. (2016). Processing of irregular polysemes in sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(11), 1798-1813. doi: 10.1037/xlm0000271.supp

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. *Behav Res Methods, 41*(4), 977-990. doi: 10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behav Res Methods, 46*(3), 904-911. doi: 10.3758/s13428-013-0403-5

Cacciari, C., & Glucksberg, S. (1991). Understanding idiomatic expressions: The contribution of word meanings. *Advances in Psychology, 77*, 217-240.

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *J Mem Lang, 27*, 668-683.

Caillies, S., & Butcher, K. (2007). Processing of Idiomatic Expressions: Evidence for a New Hybrid View. *Metaphor and Symbol, 22*(1), 79-108. doi: 10.1080/10926480709336754

Caillies, S., & Declercq, C. (2011). Kill the song-steal the show: what does distinguish predicative metaphors from decomposable idioms? *J Psycholinguist Res, 40*(3), 205-223. doi: 10.1007/s10936-010-9165-8

Cronk, B.C., Lima, S.D., & Schweigert, W.A. (1993). Idioms in sentences: Effects of frequency, literalness, and familiarity. *J Psycholinguist Res, 22*(1), 59-82.

Cutting, C.J., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Mem Cognit, 25*(1), 57-71.

Eddington, C. M. (2015). *Effects of within- and cross-language semantic ambiguity on learning and processing.* (Doctor of Philosophy), University of Pittsburgh, Pittsburgh, PA.

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

Foraker, S., & Murphy, G. L. (2012). Polysemy in Sentence Comprehension: Effects of Meaning Dominance. *J Mem Lang, 67*(4), 407-425. doi: 10.1016/j.jml.2012.07.010

Frazier, L., & Rayner, K. (1990). Taking on semantic committments: Processing multiple meanings vs. multiple senses. *J Mem Lang, 29*(2), 181-200.

Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass, 3*(1), 111-127.

Frisson, S., Harvey, D.R., & Staub, A. (2017). No prediction error cost in reading: Evidence fomr eye movements. *J Mem Lang, 95*, 200-214.

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology, 25*(6), 1366-1383.

Frisson, S., & Pickering, M. J. (2001). Obtaining a figurative interpretation of a word: Support for underspecification. *Metaphor and Symbol, 16*(3&4), 149-171. doi: 10.1080/10926488.2001.9678893

Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Mem Cognit, 8*(2), 149-156.

Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *J Mem Lang, 28,* 576-593.

Giora, R. (2002). Literal vs. figurative language: Different or equal? *Journal of Pragmatics, 34*, 487-506.

Glucksberg, S. (1991). Beyond literal meanings: The psychology of illusion. *Psychol Sci, 2*(3), 146-152. doi: 10.1111/j.1467-9280.1991.tb00122.x

Hamblin, J. L., & Gibbs, R. W. (1999). Why you can't kick the bucket as you slowly die: verbs in idiom comprehension. *J Psycholinguist Res, 28*(1), 25-39.

Holsinger, E. (2013). Representing idioms: Syntactic and contextual effects on idiom processing. *Language and Speech, 56*(3), 373-394. doi: 10.1177/0023830913484899

Jolsvai, H., McCauley, S.M., & Christiansen, M.H. (2013). Meaning overrides frequency in idiomatic and compositional multiword chunks.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behav Res Methods, 42*(3), 627-633. doi: 10.3758/BRM.42.3.627

Klepousniotou, E. (2002). The Processing of Lexical Ambiguity: Homonymy and Polysemy in the Mental Lexicon. *Brain Lang, 81*(1-3), 205-223. doi: 10.1006/brln.2001.2518

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *J Neurolinguistics, 20*(1), 1-24. doi: 10.1016/j.jneuroling.2006.02.001

Klepousniotou, E., Pike, G. B., Steinhauer, K., & Gracco, V. (2012). Not all ambiguous words are created equal: an EEG investigation of homonymy and polysemy. *Brain Lang, 123*(1), 11-21. doi: 10.1016/j.bandl.2012.06.007

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: the comprehension of polysemy depends on sense overlap. *J Exp Psychol Learn Mem Cogn, 34*(6), 1534-1543. doi: 10.1037/a0013012

Konopka, A. E., & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology, 58*(1), 68-101. doi: 10.1016/j.cogpsych.2008.05.002

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: tests in linear mixed effects models. R package version 2.0-20. *URL: https://cran. rproject. org/web/packages/lmerTest. Accessed, 15*, 2016.

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Mem Cognit, 36*(6), 1103-1121. doi: 10.3758/MC.36.6.1103

Nordmann, E., Cleland, A. A., & Bull, R. (2014). Familiarity breeds dissent: Reliability analyses for British-English idioms on measures of familiarity, meaning, literality, and decomposability. *Acta Psychol (Amst), 149*, 87-95. doi: 10.1016/j.actpsy.2014.03.009

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy, 3*(2), 143-184.

Nunberg, G., Sag, I.A., & Wasow, T. (1994). Idioms. *Language, 70*(3), 491-538.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: the role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 776-784. doi: 10.1037/a0022531

Smolka, E., Rabanus, S., & Rösler, F. (2007). Processing Verbs in German Idioms: Evidence Against the Configuration Hypothesis. *Metaphor and Symbol, 22*(3), 213-231. doi: 10.1080/10926480701357638

Spears, R. (Ed.) (2007) McGraw-Hill's dictionary of American idioms and phrasal verbs. McGraw-Hill.

Swinney, D.A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior, 18*, 523-534.

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Mem Cognit, 37*(4), 529-540. doi: 10.3758/MC.37.4.529

Titone, D., & Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics, 31*, 1655-1674.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning, 61*(2), 569-613. doi: 10.1111/j.1467-9922.2010.00622.x

Warren, T., Milburn, E., Patson, N. D., & Dickey, M. W. (2015). Comprehending the impossible: what role do selectional restriction violations play? *Language, Cognition and Neuroscience, 30*(8), 932-939. doi: 10.1080/23273798.2015.1047458

Weiss, A.F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (2017). Comprehension demands modulate re-reading, but not first pass reading behavior. *Quarterly Journal of Experimental Psychology*, 1-37.