Idioms show effects of meaning relatedness and dominance similar to those seen for

ambiguous words

Evelyn Milburn[1] and Tessa Warren[2]

Norwegian University of Science and Technology[1]

University of Pittsburgh[2]

Correspondence to:
Evelyn Milburn
NTNU Dragvoll
7049 Trondheim
Trøndelag, Norway
Email: evelyn.a.milburn@ntnu.no
Phone: +47 462 60 140

Abstract

Does the language comprehension system resolve ambiguities for single and multiple word units similarly? We investigate this question by examining whether two constructs with robust effects on ambiguous word processing—meaning relatedness and meaning dominance—have similar influences on idiom processing. Eye tracking showed that: (1) idioms with more related figurative and literal meanings were read faster, paralleling findings for ambiguous words (Eddington & Tokowicz, 2015), and (2) meaning relatedness and meaning dominance interacted to drive eye movements on idioms just as they do on polysemous ambiguous words (Foraker & Murphy, 2012). These findings are consistent with a language comprehension system that resolves ambiguities similarly regardless of literality or the number of words in the unit.

Keywords: figurative language, ambiguity, idioms, ambiguous words, eye movements and reading

INTRODUCTION

Language is figurative when its intended meaning is not a simple combination of the relevant words' meanings (Glucksberg, 1991). Considerable research has focused on the differences between literal and figurative language processing (e.g. Gibbs, 1980), but more recent research has shown parallels between them (Cutting & Bock, 1997; Giora, 2002; Konopka & Bock, 2009). The present research builds from the assumption that the language processing system processes flexibly across literal and figurative single words and multiword chunks, to test whether the factors affecting lexical ambiguity resolution similarly affect idiom ambiguity resolution.

Historically, idioms have been assumed to be represented and processed as multi-word units (Gibbs, 1980; Swinney & Cutler, 1979), but now it is thought that the same may hold for common literal strings (or *lexical bundles*). More frequent literal strings are processed more quickly (e.g. Arnon & Snider, 2010) and remembered more accurately (Tremblay, Derwing, Libben, & Westbury, 2011) than less frequent ones. Indeed, findings of facilitated processing for non-adjacent frequent collocations (Vilkaite, 2016) suggest that the language system may process words and bundles simultaneously. This is consistent with classic findings suggesting that idioms are not *only* processed as units —particularly the fact that changing a word in an idiom does not always disrupt processing (e.g. Hamblin & Gibbs, 1999; Nordmann, Cleland, & Bull, 2014).

Ambiguity resolution is important in idiom processing; most idioms have both literal and figurative meanings (ex. *on the fence*) (Cronk, Lima, & Schweigert, 1993). Several models propose that idiom comprehension involves compositional analysis of the literal meaning and simultaneous direct retrieval of the figurative meaning (for an overview, see Cacciari, 2014), consistent with a processing framework in which single words and lexical bundles are processed simultaneously. Sprenger et al.'s Superlemma Hypothesis (2006)

assumes that words and idioms are both represented by single units—for idioms, the titular *superlemmas*—at both lemma and concept levels. Connections between individual-word lemmas and the idiom's superlemma account for priming effects wherein one word from an idiom primes the whole idiom and vice versa (Sprenger et al., 2006).

Under this view, the same factors that affect lexical ambiguity resolution should affect idiom ambiguity resolution. Meaning relatedness is one construct that appears to have such parallel effects. Homonyms, like *bank*, have multiple unrelated meanings. This low meaning relatedness sometimes disadvantages homonym processing, possibly because the meanings compete for activation (Armstrong & Plaut, 2016). Polysemes, like *sheet*, have multiple related meanings (referred to as senses). Their high relatedness generally aids polyseme processing (Eddington & Tokowicz, 2015). Similarly, higher relatedness between idioms' literal and figurative meanings usually facilitates processing (Caillies & Butcher, 2007; Gibbs et al., 1989), particularly when semantic processing is heavily recruited (Libben & Titone, 2008). *Decomposability* is an idiom-specific meaning relatedness construct indexing the semantic relationship between the meanings of the individual words in an idiom and the idiom itself (e.g. Gibbs, Nayak, Bolton, & Keppel, 1989; Caillies & Butcher, 2007). Titone and Connine (1999) argued that decomposability should affect idiom processing the same way that meaning relatedness affects lexical processing: for decomposable idioms, the meanings' relatedness should facilitate their joint activation, but for nondecomposable idioms, the unrelated meanings should compete. Indeed, Titone and Connine (1999) observed processing difficulty for nondecomposable, but not decomposable, idioms.

Another construct with similar effects on both idiom and ambiguous word processing is meaning dominance: the relative frequency of one meaning over others (Duffy, Morris, & Rayner, 1988). Figurative idiom meanings are frequently dominant over literal meanings (e.g. Cronk, Lima, & Schweigert, 1993). However, the appropriateness of a particular idiom

meaning can be clarified and guided by context (Fanari, Cacciari, & Tabossi, 2010), and mismatches between contextual expectations and meaning dominance can result in processing disruption (Holsinger & Kaiser, 2013).

Studies of ambiguous word comprehension suggest that effects of meaning dominance and meaning relatedness may interact, although multiple interactive patterns have been observed. Brocher and colleagues (2018) crossed the meaning dominance and meaning relatedness of ambiguous words that were followed by a clause that biased their subordinate meaning. Reading times on the disambiguating region were longer following words with less related meanings when one meaning was strongly dominant, suggesting that a word's dominant meaning may interfere with the processing of its less-dominant meaning when they are unrelated. On the other hand, Foraker and Murphy (2012) embedded ambiguous words with related meanings into sentences with both biasing contexts and disambiguating final clauses. They observed processing slowdowns for words with one strongly dominant sense, but no corresponding slowdowns for words with more balanced senses, regardless of context or disambiguation. The findings of these studies are contradictory: both found processing difficulty when one meaning was strongly dominant, but the studies differ on whether concurrent strong meaning relatedness aids or impedes processing. If ambiguous words and idioms are processed similarly, either of these interactions may appear for idioms.

The present research aims to determine whether patterns of meaning dominance and meaning relatedness observed in ambiguous word research generalize to idiom processing. Following Brocher and colleagues (2016; 2018) and Foraker and Murphy (2012), we embedded idioms varying in meaning relatedness and figurative meaning dominance into contexts biasing either their literal or figurative interpretations, and tracked participants' eyes as they read them. Whereas these previous studies included postcritical regions disambiguating their ambiguous words, in the current study all material after the idiom was

consistent with both of its interpretations. Therefore any observed effects, even in re-reading behaviors, should reflect relatively undisrupted processing rather than revision processes. Critically, these previous studies all found effects on their critical words; we therefore expect to see effects on the idioms in this study.

If the language processing system uses the same mechanisms to resolve ambiguities regardless of literality or the number of words in the unit, we expect to see similar effects of meaning dominance and relatedness during idiom resolution as have previously been seen for ambiguous words. In particular, we predict that idioms with more related figurative and literal meanings should be read more quickly and re-read less, congruent with previous studies (e.g. Klepousniotou, 2002; Titone & Connine, 1999). If biasing context affects ambiguous word and idiom processing similarly, then idioms with highly dominant figurative meanings should be read more slowly in literally-biasing contexts, following Brocher et al.'s (2016) findings with ambiguous words. Finally, we predict that meaning dominance and meaning relatedness will interact to drive idiom comprehension, although it is not clear whether to expect the interaction from Brocher et al. (2018) or Foraker and Murphy (2012).

METHODS

Participants

Thirty-six undergraduate native English speakers (who did not participate in the norming described below) participated.

Materials

Participants read three-sentence passages while their eyes were tracked. Passages consisted of a biasing context sentence, an idiom sentence, and a wrap-up sentence. Context sentences biased either the less dominant literal or more dominant figurative interpretation of the idiom. There were 45 items with two conditions each:

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

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

Idiom sentences contained a neutral precritical region (italicized), an idiom (underlined), and a neutral postcritical region (bracketed). Idiom and wrap-up sentences were consistent with both interpretations of the idiom. Context-bias conditions were counterbalanced between participants. Each sentence appeared on its own line, and the critical idiom region was always in the middle of the second line.

We included 80 filler passages of varying lengths and structures; some contained figurative language. After 40 of the filler passages a yes/no comprehension question appeared, half of which required a "yes" response.

Norming

Figurative dominance, meaning relatedness, and idiom familiarity norm values (Table 1) were used as continuous predictors in statistical analyses.

Table 1. Idiom figurative dominance, meaning relatedness, and familiarity descriptive statistics

Norm	Mean	SD	Range
Figurative Dominance (50-100%)	74.89	9.85	50.17 - 89.64
Meaning Relatedness (1-7)	3.60	.91	1.14 – 5.25
Idiom Familiarity (1-7)	5.17	.52	4.08 – 5.92

Figurative Dominance

Fifty-seven participants normed 140 idioms. Participants saw each idiom with a slider between its literal and figurative meanings. They moved the slider to indicate what percentage of the time they expected the idiom to have each meaning.

Meaning Relatedness

The same 57 participants saw each idiom's literal and figurative meanings and rated how related they were on a scale of 1 (*very unrelated*) to 7 (*very related*). Participants were not told that the phrases they saw were idiom meanings. We did not collect finer-grained data regarding *why* participants rated idioms the way they did: high relatedness could therefore have indexed metaphorical correspondences between individual words and aspects of the figurative meaning (Gibbs et al., 1989), general contributions of single words to figurative meaning (Caillies, & Butcher, 2007), the ease with which the motivation for the idiom's figurative use could be understood or invented (Nunberg, Sag, & Wasow, 1994), or something idiosyncratic to each participant. We chose our approach because it is most similar to how meaning relatedness is usually normed in word ambiguity studies (e.g. Klepousniotou et al., 2008). Note additionally that ratings of idiom decomposability or transparency tend to be unstable (e.g. Keysar & Bly, 1995); our approach was intended to minimize this instability while maximizing parallels to ambiguous word norming.

Participants did not rate the same idioms on both dominance and relatedness. Table 2 shows example idioms with higher and lower figurative dominance and meaning relatedness values.

Idiom Familiarity

Fifty-three participants rated 240 idioms on a scale of 1 (*very unfamiliar*) to 7 (*very familiar*).

Table 2. Example idioms with higher and lower figurative dominance/meaning relatedness values

Meaning Relatedness	Figurative Dominance	Idiom	Relatedness Score	Dominance Score
High	Biased	on the fence	4.75	89%
	Balanced	deliver the goods	4.3	59%
Low	Biased	fall off the wagon	2.71	78%
	Balanced	cut down to size	2.08	61%

Idiom Selection Criteria

The 45 selected idioms had comprehensible literal interpretations (ensuring meaning ambiguity), scored between 4 and 7 in the familiarity norm (ensuring accessibility of figurative meanings), and had dominant (>50%) figurative meanings. Lower figurative dominance scores meant that the rated likelihoods of literal and figurative meanings were relatively balanced.

Context Bias

Twenty-four participants saw each passage truncated after the idiom, and rated on a scale of 1 (*figurative*) to 7 (*literal*) which meaning of the idiom was appropriate. Participants preferred the figurative meaning after the figuratively-biased context (M = 1.55; SD = .42) and the literal meaning after the literally-biased context (M = 5.85; SD = .85; paired-sample *t*-test: $t(44) = -30.77$; p<.05).

Progressive Naturalness

Six participants rated the naturalness of each passage truncated before the idiom, after the idiom, after the idiom sentence, and after the full item, on a scale of 1 (*very natural*) to 7 (*very unnatural*); see Table 3. We used naturalness ratings as predictors in analyses.

Table 3. Progressive naturalness norm descriptive statistics

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

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. Three characters equaled 1º of visual angle. The experiment lasted approximately 30 minutes. Chin and forehead rests minimized head movements. The tracker was calibrated using a 9-point calibration grid; additional re-calibration was performed as necessary. A single-point drift check was performed every five trials.

RESULTS

We analyzed four 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 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. We focused on eye-movement measures that involved re-reading because these are often where meaning-related effects appear (e.g. Frisson & Pickering, 2001), and we wanted to minimize the number of statistical tests. First fixation was included to capture any early spillover effects in the postcritical region.

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 longer fixation were combined. Other fixations shorter than 60 ms or longer than 1500 ms were removed (Brocher et al., 2016). Three trials in the literal context-bias condition were removed due to calibration errors. After processing, 1,561 trials were analyzed.

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, 2017; ver. 2.0-20). Models were fit using the fullest random effects structure that would allow convergence (Barr, Levy, Scheepers, & Tily, 2013). Models included fixed effects of context bias (literal vs. figurative), figurative dominance, meaning relatedness, and idiom familiarity. 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, and random slopes of progressive naturalness score within items and figurative dominance, meaning relatedness, idiom familiarity, and progressive naturalness score within participants. To improve interpretability, we reverse-coded progressive naturalness. Note that meaning relatedness and figurative dominance were treated as continuous predictors in statistical models. Table 4 shows means for all eye-movement measures in all regions. Because there were no effects of context type in statistical analyses, means are shown collapsed over context bias.

Table 4. Descriptive statistics for all eye movement measures in all analysis regions, for high and low meaning relatedness and biased and balanced figurative dominance, collapsed over context type

Region		Measure	Meaning Relatedness	Figurative Dominance	Mean (ms)	SD

Precritical	First Fixation	High	Biased	212.91	72.73
			Balanced	208.12	70.38
		Low	Biased	215.67	87.97
			Balanced	209.66	81.76
	Go Past	High	Biased	535.51	366.02
			Balanced	447.25	295.62
		Low	Biased	488.74	331.81
			Balanced	472.15	374.23
	Rereading	High	Biased	340.65	259.97
			Balanced	300.55	190.51
		Low	Biased	288.19	173.30
			Balanced	325.97	238.32
	Total Time	High	Biased	586.48	405.99
			Balanced	500.33	332.69
		Low	Biased	559.79	416.46
			Balanced	528.59	380.06
Critical	First Fixation	High	Biased	215.00	72.52
			Balanced	210.01	63.83
		Low	Biased	216.96	69.61
			Balanced	208.45	65.14
	Go Past	High	Biased	632.11	337.02

			Balanced	524.60	246.75
		Low	Biased	562.28	252.07
			Balanced	590.09	334.77
	Rereading	High	Biased	331.74	188.67
			Balanced	308.93	154.41
		Low	Biased	332.48	202.09
			Balanced	362.05	238.71
	Total Time	High	Biased	694.13	340.84
			Balanced	606.72	320.94
		Low	Biased	641.48	305.88
			Balanced	654.24	390.14
Postcritical	First Fixation	High	Biased	235.77	90.91
			Balanced	226.90	71.68
		Low	Biased	227.95	85.50
			Balanced	220.16	83.98
	Go Past	High	Biased	800.03	604.60
			Balanced	811.82	588.73
		Low	Biased	760.32	593.05
			Balanced	772.03	576.50
	Rereading	High	Biased	394.27	343.87
			Balanced	415.62	417.60
		Low	Biased	482.60	386.00
			Balanced	425.02	363.95

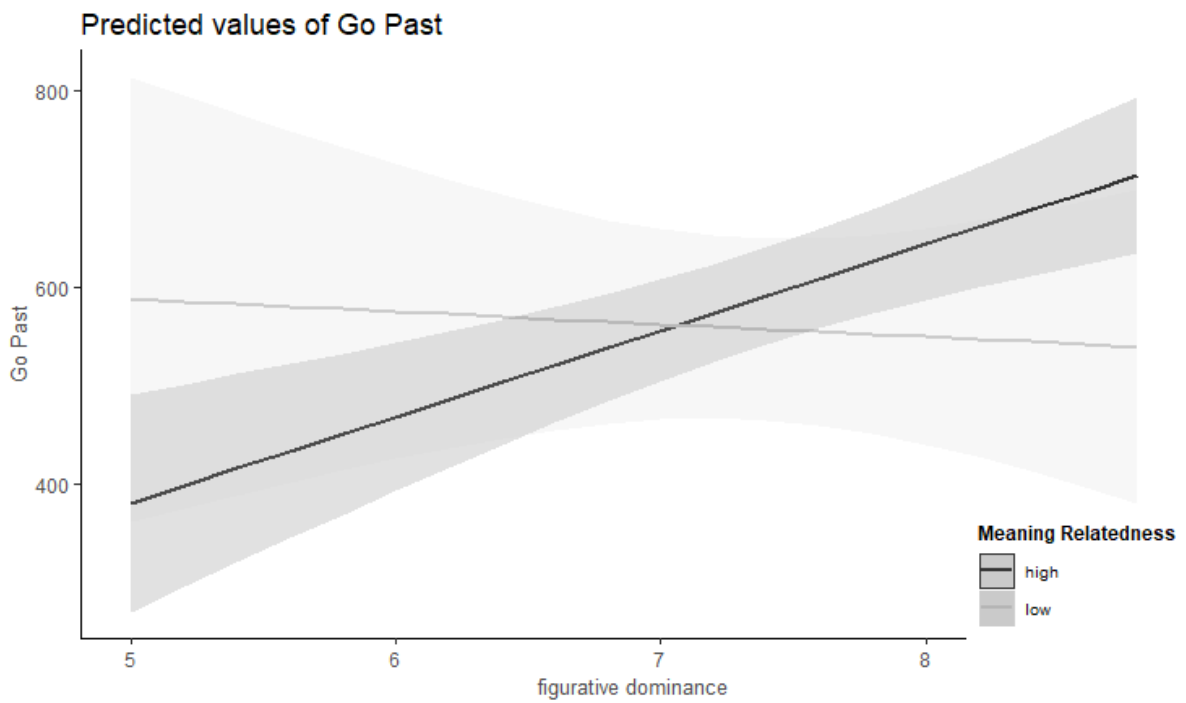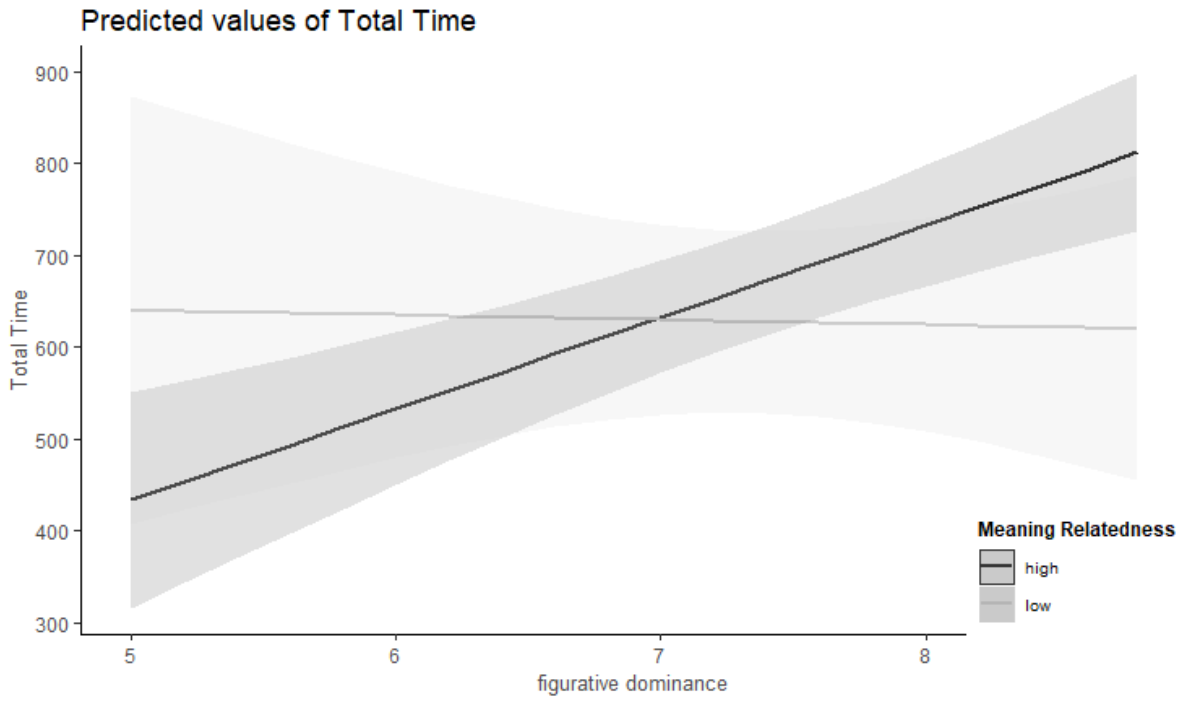|	Total				
	Time	High	Biased	731.16	472.94
			Balanced	731.96	418.10
		Low	Biased	699.57	446.78
			Balanced	698.98	429.63

Pre-Critical Region

　　　　There were no reliable effects.

Critical Region

　　　　Models for go past and re-reading converged with full random effects structure. For first fixation, random slopes of meaning relatedness and figurative dominance within participants and progressive naturalness score within items were dropped. For total time, the random slope of progressive naturalness score within items was dropped. First fixation showed no reliable effects. Go past, re-reading, and total time all showed a reliable effect of figurative dominance: as it increased, reading time decreased (Go Past: β=-254.06; SE=87.97; t=-2.89; p<.05; Re-reading: β=-91.86; SE=45.61; t=-2.01; p=.051; Total time: β=-214.91; SE=91.01; t=-2.36; p<.05). Go past and total time also showed reliable effects of meaning relatedness: as relatedness increased, reading time decreased (Go Past: β=-547.01.42; SE=170.21.; t=-3.21; p<.05; Total time: (β=-465.92; SE=176.08; t=-2.65; p<.05).). In all three measures, significant main effects were qualified by a significant interaction between figurative dominance and meaning relatedness (Go Past: β=76.96; SE=23.05; t=3.28; p<.05; Re-reading: β=26.60; SE=12.24; t=2.17; p<.05; Total time: β=65.65; SE=24.31; t=2.70; p<.05): ==increased meaning relatedness slowed processing when figurative dominance was high but speeded processing when figurative dominance was low== (==Figure 1==). Finally, as familiarity increased, reading time decreased in all three measures (Go

past: β=-71.54; SE=39.35; t=-1.82; p=.08; Re-reading: β=-38.93; SE=20.33; t=-1.92; p=.06;

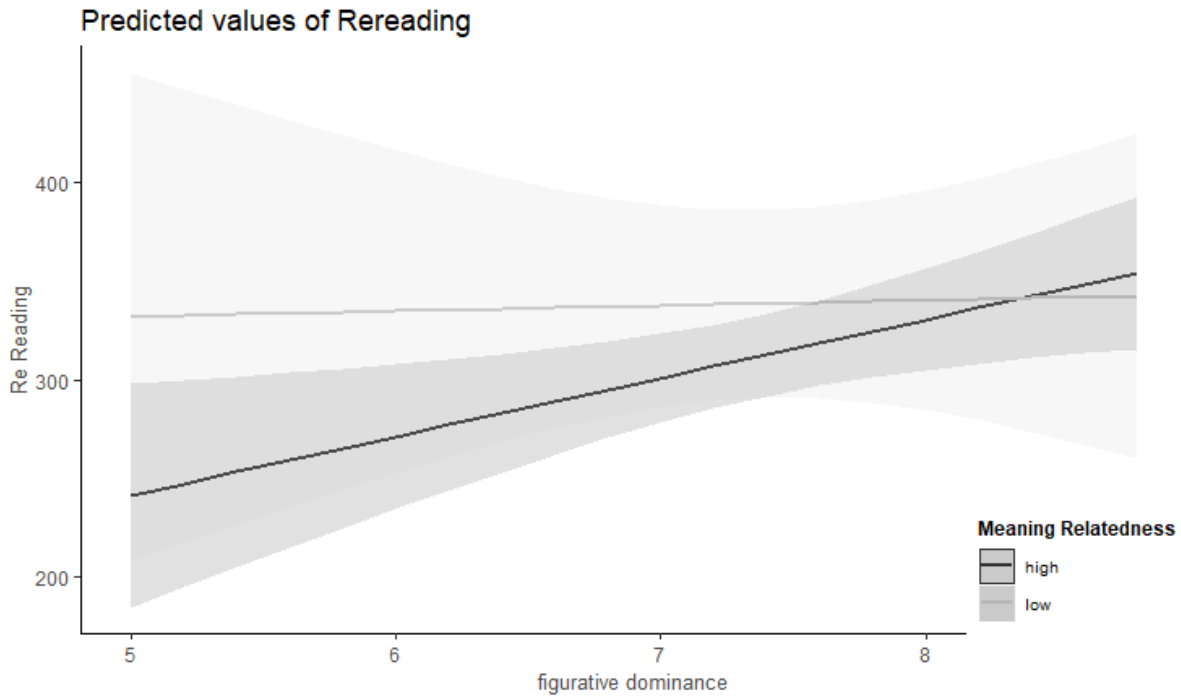Total time: β=-91.72; SE=40.71; t=-2.25; p<.05).

Figure 1. Estimated effects of figurative dominance and meaning relatedness on Total Time, Go Past, and Re-Reading in the critical region, collapsed over context bias. X-axes show average rated percentage of time (divided by 10) that participants would expect the string to have the figurative meaning Note that although meaning relatedness was treated as a continuous variable in analyses, it is presented here using a median split for ease of interpretation.

Post-Critical Region

Models for go past and re-reading converged with full random effects structure. For first fixation and total time, the random slope of progressive naturalness score within items was dropped. First fixation showed a marginal interaction between context bias and figurative dominance ($\hat{\beta}$=35.29; SE=18.81; t=1.88; p=.06) and a marginal interaction between context bias, figurative dominance, and meaning relatedness ($\hat{\beta}$=-9.06; SE=5.02; t=-1.81; p=.07). Go past and total time both showed a reliable effect of naturalness: as it increased, reading time decreased (Go past: $\hat{\beta}$=-75.81; SE=25.00; t=-3.03; p<.05; Total time: $\hat{\beta}$=-39.47; SE=17.62; t=-2.24; p<.05)

DISCUSSION

The findings of this study suggest parallels between ambiguous idiom and ambiguous word resolution processes. First, more figuratively-dominant idioms were read faster than idioms with more balanced meanings, similar to ambiguous words (Duffy et al., 1988). Second, idioms with more related literal and figurative meanings were read faster and re-read less. Third, meaning relatedness and figurative dominance interacted, paralleling post-hoc results from Foraker and Murphy (2012) showing processing disadvantages for polysemes with highly related but biased meanings

Contrary to Brocher et al. (2018)'s ambiguous word findings, we observed no processing disruption when a figuratively-dominant idiom appeared within a literally-biased context. Context only played a role in a marginal interaction between context bias, figurative dominance, and meaning relatedness on first fixation duration in the postcritical region, and it is unclear what mechanism this interaction might represent. This lack of context effects is not unprecedented; Conklin and Schmitt (2008) observed none on the interpretation of ambiguous formulaic phrases. The current lack of context effects may have arisen because the neutral post-idiom regions did not force participants to fully disambiguate the idioms. When context effects appear, they are often on disambiguating regions following the ambiguous unit (see Titone & Connine, 1999; Foraker & Murphy, 2012). Another possible explanation for the lack of context effects is that both literal and figurative contexts produced processing slowdowns for biased idioms with related meanings, although for different reasons. Slowdowns in figurative contexts could be due to strong enrichment of the dominant sense, whereas slowdowns in literal contexts could be due to competition between the subordinate context and the dominant figurative meaning.

Meaning relatedness and figurative dominance interacted such that, in the critical region, high meaning relatedness slowed processing when figurative dominance was high but speeded processing when figurative dominance was low. This closely parallels Foraker and

Murphy (2012)'s interaction, in which high sense dominance and high relatedness disadvantaged polyseme processing, and is the opposite of Brocher et al (2018)'s interaction across polysemes and homonyms, in which high sense dominance and *low* relatedness conferred disadvantages. Foraker and Murphy hypothesized that the strong dominance of one sense may increase between-sense competition, over-riding the advantage normally afforded by high sense relatedness. Perhaps when one sense is strongly dominant, it becomes more differentiated from other senses, resulting in a polyseme that is processed more like a homonym—although this explanation neglects the role of context during processing. The current interaction could indicate that our participants processed the highly familiar idioms tested here more like polysemes than homonyms. Indeed, large-scale norming studies show that idiom familiarity and perceived compositionality (a measure related to meaning relatedness) are positively correlated (Nordmann et al. 2014). On the other hand, our norming indicated that some of our idioms had strongly unrelated meanings. Ultimately, more work will be needed to understand the mechanisms driving these interactions.

The observed parallels between lexical and idiom ambiguity resolution open new questions about idiom representation. Research into lexical ambiguity resolution has begun to investigate: (1) whether different lexical meanings are stored separately or together, (2) if together, whether the representation is underspecified or based on core features, and (3) whether the answers to these questions vary according to the type of ambiguous word (Eddington & Tokowicz, 2015). Research into idiom ambiguity resolution has barely begun to grapple with these questions, which become even more relevant if some idioms might be lexical bundles with stored literal and figurative meanings. The Superlemma Hypothesis, in which idioms and single words have similar underlying representational structures, seems a good framework within which to start considering these questions.

The current findings are consistent with a language processing system that resolves ambiguities similarly regardless of literality or the number of words in the unit. We found that meaning dominance and meaning relatedness in idiom ambiguity resolution interacted such that higher relatedness and figurative dominance elicited processing slowdowns, paralleling previous findings for ambiguous words (Foraker & Murphy, 2012). If—as suggested by these results—the same representations and mechanisms are in play for resolution of both idioms and ambiguous words, then representational questions that are specific to either literature may be usefully applied to both.

Works Cited

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, *31*(7), 940–966. https://doi.org/10.1080/23273798.2016.1171366

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82. https://doi.org/10.1016/j.jml.2009.09.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

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

Cacciari, C., & Tabossi, P. (2014). Idioms: Processing, structure, and interpretation. Psychology Press.

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

Cronk, B. C., Lima, S. D., & Schweigert, W. A. (1993). Idioms in sentences: Effects of frequency, literalness, and familiarity. *Journal of Psycholinguistic Research*, *22*(1), 59–82. https://doi.org/10.1007/BF01068157

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

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*, 429–446.

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

Fanari, R., Cacciari, C., & Tabossi, P. (2010). The role of idiom length and context in spoken idiom comprehension. *European Journal of Cognitive Psychology*, *22*(3), 321–334. https://doi.org/10.1080/09541440902843866

Foraker, S., & Murphy, G. L. (2012). Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of Memory and Language*, *67*(4), 407–425. https://doi.org/10.1016/j.jml.2012.07.010

Frisson, S. (2009). Semantic Underspecification in Language Processing. *Language and Linguistics Compass*, *3*(1), 111–127. https://doi.org/10.1111/j.1749-818X.2008.00104.x

Frisson, S., & Pickering, M. J. (1999). The Processing of Metonymy: Evidence From Eye Movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *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. https://doi.org/10.1080/10926488.2001.9678893

Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, *8*(2), 149–156. https://doi.org/10.3758/BF03213418

Gibbs, R. W., Nayak, N. P., Bolton, J. L., & Keppel, M. E. (1989). Speakers' assumptions about the lexical flexibility of idioms. *Memory & Cognition*, *17*(1), 58–68. https://doi.org/10.3758/BF03199557

Giora, R. (2002). Literal vs. figurative language: Different or equal? *Journal of Pragmatics*, *34*(4), 487–506. https://doi.org/10.1016/S0378-2166(01)00045-5

Glucksberg, S. (1991). Beyond Literal Meanings: The Psychology of Allusion. *Psychological Science*, *2*(3), 146–152. https://doi.org/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, *28*(1), 15. https://doi.org/10.1177/0023830913484899

Holsinger, E., & Kaiser, E. (2013). Processing (non)compositional expressions: Mistakes and recovery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 866–878. https://doi.org/10.1037/a0030410

Keysar, B., & Bly, B. (1995). Intuitions of the Transparency of Idioms: Can One Keep a Secret by Spilling the Beans? *Journal of Memory and Language*, *34*(1), 89–109. https://doi.org/10.1006/jmla.1995.1005

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

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, *36*(6), 1103–1121. https://doi.org/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 Psychologica*, *149*, 87–95. https://doi.org/10.1016/j.actpsy.2014.03.009

Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language*, *70*(3), 491. https://doi.org/10.2307/416483

R Core Team. (2013). R: A language and environment for statistic computing (Version 3.0.1). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, *18*(5), 523–534. https://doi.org/10.1016/S0022-5371(79)90284-6

Titone, D. A., & Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics*, *31*(12), 1655–1674. https://doi.org/10.1016/S0378-2166(99)00008-9

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing Advantages of Lexical Bundles: Evidence From Self-Paced Reading and Sentence Recall Tasks: Lexical Bundle Processing. *Language Learning*, *61*(2), 569–613. https://doi.org/10.1111/j.1467-9922.2010.00622.x

Vilkaite, L. (2016). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1632–1642.