



How discourse constraints influence neurolinguistic mechanisms during the comprehension of proverbs

Todd R. Ferretti¹ · Albert N. Katz² · Christopher A. Schwint¹ · Courtney Patterson¹ · Dagna Pradzynski¹

Published online: 8 May 2020
© The Psychonomic Society, Inc. 2020

Abstract

We use event-related brain potential (ERP) methodology to examine the influence of the linguistic markers *literally speaking* and *figuratively speaking* on the comprehension of proverbs (e.g., *The cat is out of the bag*). Our results show that slow cortical potentials at anterior electrode sites varied in amplitude across the proverbs as a function of the presence or absence of the markers, the presence and absence of discourse contexts, and the familiarity of the proverbs. The results demonstrate that the integration of literal meaning into context is easier than figurative meaning, and argues against models of figurative language processing that hold that comprehenders are obligated either to first process the literal or figurative sense of the trope.

Keywords Erp · Discourse · Proverbs

The classic studies on the processing of figurative language employed behavioral indices, such as reading time, to examine comprehension of tropes presented without an accompanying discourse context and found that, in general, it took longer to process figurative than comparable literal sentences. In subsequent years, researchers examined how and when people comprehend figurative statements when they are placed in discourse contexts. In general, a different pattern of findings emerge when figurative language is placed in discourse; most often with a rich and elaborated context, there are no differences observed in the behavioral indices, especially when familiar or salient non-literal language is employed (Katz, 1966).

These findings, especially the elimination of processing differences for familiar tropes placed in discourse, have direct implications for extant models of figurative language processing. The general logic has been to assume that the discourse context sets up an interpretive framework that provides information useful for integrating the upcoming trope, with different theories making explicit predictions on whether or not that information would facilitate comprehension at the earliest moments of comprehension. For example, the standard pragmatic

model posits that figurative statements are constructed only after an initial and obligatory literal interpretation is determined to be inconsistent with the preceding context, and that additional processing mechanisms are involved in figurative language comprehension than during the comprehension of literal language (Grice, 1975; Searle, 1979). The graded salience model similarly holds that the important distinction is not literal or figurative language but rather whether the language is conventional (salient) or nonconventional (nonsalient meaning); there is obligatory processing regardless of the nature of the preceding discourse context but now one is obligated to process the most salient (i.e., familiar or conventional) meaning of a statement, be it figurative or literal (Giora, 2003). Context can boost activation of non-salient meanings but never at a cost to the activation of the salient meaning. Thus, in this model, integrating the salient meaning of highly familiar nonliteral statements, such as the proverb *Don't count your chickens before they hatch*, into discourse contexts that support the proverbial meaning should always be at least as easy, if not easier, than when the same sentence is presented in contexts biased toward the less salient meaning (i.e., literal) of the proverbs.

In contrast, there are theories that do not posit obligatory processing of the figurative statement, even when the statement is highly familiar, but argue instead that the initial processing of the statement could be consistent with either a literal or nonliteral reading, depending on the nature of the preceding context. Gibbs (1994) for instance posits that with a sufficiently rich, appropriate, and elaborated context the non-literal sense of a figurative statement could be primed and

✉ Todd R. Ferretti
tferretti@wlu.ca

¹ Centre for Cognitive Neuroscience and Department of Psychology, Wilfrid Laurier University, 75 University Avenue, Waterloo, Ontario N2L 3C5, Canada

² Western University, London, Ontario, Canada

hence as easy to integrate as the literal sense (when placed into a sufficiently rich, appropriate, and elaborated literal context), even for nonconventional usages. Gibbs's approach to how context can influence the activation of both literal and figurative meanings of statements is consistent with constraint-based models of language processing. A constraint-based approach to figurative language focuses on the different constraints present in the context that may increase or decrease the activation of figurative meanings relative to literal meanings during language processing. One of the main assumptions of this approach, which follows directly from constraint-based models in general (McDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998), is that the determination of the activation of literal and figurative meaning of a figurative statement is a competitive process, and the meaning that is most activated is determined by the strength of the different sources of information (e.g., lexical, syntactic, conceptual, pragmatic) that support the competing meanings. In principle, either the literal sense or the figurative sense of a figurative statement might be activated. Thus, with constraint satisfaction models the emphasis is on examining the nature and strength of contextual information supporting a literal or figurative interpretation.

Research that has attempted to tease apart these various theories have tended to use behavioral indices of comprehension, such as reading time, which may not be as sensitive to picking up processing differences as are measures based on cortical brain potentials, as will be employed here. Moreover, extant studies have tended to manipulate discourse context such that either the literal or figurative sense of a target statement is supported. Much less consideration has been given to other aspects of the experimental contexts that are produced. One such aspect of contexts is the degree of semantic overlap shared by the context and a target sentence. Specifically, some researchers argue that the different conceptual domains relevant for comprehending a statement tend to be more distant for figurative than for literal statements. Because more distant domains have less conceptual overlap with a target statement than do less distant domains, the resulting differences in overlap should lead to greater conceptual integration difficulty for figurative statements (Blasko, 1999; Coulson & Van Petten, 2002). Consider for instance constructing discourse contexts for the familiar proverb *Don't count your chickens before they hatch*. Proverbs are distinguished relative to other forms of figurative language (e.g., metaphor, irony), because their meanings tend to be valid in both literal and figurative contexts (i.e., you really shouldn't count your chickens before they hatch). In a literal-supporting context, the discourse would be conceptually coherent and probably consist of words, such as *chicken* or *birth*, whereas in figurative-supporting contexts the topic would not be about *chickens* or *hatching* per se. Thus, on processing the proverb the comprehender would have to both understand the figurative

sense of the proverb and integrate it into information made available by the preceding context, which involves associating the new topic (*chickens*) with the existing discourse structure built around another more distant topic. One purpose of the present paper is to examine whether lexical-semantic relations between words in a proverb and those in the discourse context that precedes the proverb facilitates integration of meaning.

As noted above, most of the existing literature on figurative language processing employs behavioral data. Research using reading time and reaction time as dependent measures have challenged the validity of the standard pragmatic approach (Grice, 1975; Searle, 1979), because many examples of research show that it does not necessarily take longer to construct figurative than literal meaning (Gibbs, Bogdanovich, Sykes, & Barr, 1997; Katz & Ferretti, 2001). For example, Katz and Ferretti (2001) recorded word-by-word reading times while people read familiar proverbs (e.g., *Don't count your chickens before they hatch*) and unfamiliar proverbs (e.g., *An empty sack cannot stand upright*) in rich discourse contexts that were biased toward either the literal or figurative meaning of the statements. Their results showed that people read the familiar proverbs at the same rate throughout the proverbs, but, for unfamiliar proverbs, people read the unfamiliar proverbs much more quickly in literal than figurative contexts. Moreover, these reading time differences appeared within the first few words of the proverbs, findings inconsistent with the standard pragmatic approach. These results, however, are consistent with the graded salience model as this approach holds that context-driven facilitation of less salient meaning cannot be done at a cost to the access of salient meaning (see Giora, 2003). Thus, according to this approach, it is possible for familiar proverbs to be read similarly in literal and figurative contexts. Our reading of this position is that it makes a further assertion: it should never be the case that for familiar proverbs the less salient literal sense of the proverbs is easier to integrate into context than the salient figurative sense.

In addition to the problems these results posit for the standard pragmatic approach, Katz and Ferretti's (2001) results also are problematic for approaches that highlight the combinatorial processes between words and contexts (Coulson & Van Petten, 2002; Gentner & Wolff, 1997). Such theories should predict that familiar proverbs would be read more quickly in literal contexts, because, unlike metaphors, literal contexts often are consistent with both the literal and figurative meaning of proverbs (*You really shouldn't count your chickens before they hatch*). Because of this general property of proverbs, there is typically greater conceptual overlap between the content words in the literally biasing contexts and the words in the proverbs. Thus, the integration of the proverbial sense of the statement into figurative discourse contexts should be more difficult than the integration of the literal sense because figurative contexts will not be as biasing as the literal contexts.

Neurophysiology of Figurative Language Processing

More recently, researchers have used event-related brain potential methodology (ERP) to investigate how and when people interpret figurative language (Bianchi, Shalom, & Kamienkowski, 2019; Canal, Pesciarelli, Vespignani, Molinaro, & Cacciari, 2017; Coulson & Van Petten, 2002; Ferretti, Schwint, & Katz, 2007b; Katz, Blasko, & Kazmerski, 2004; Laurent, Denhières, Passerieux, Iakimova, & Hardy-Baylé, 2006; Pynte, Besson, Robichon, & Poli, 1996). This research has typically concentrated on single word indices of text integration difficulty (e.g., N400 and Late Positivity) that either completed the sentence in a figurative manner (usually a metaphor or idiom) or literal manner. The N400 is the most widely used ERP measure for investigating the semantic integration of words in text. Between 300–500 ms after stimulus onset, words that are easier to integrate produce less negativity, particularly at central and posterior head locations (Kutas & Hillyard, 1980; Kutas, Van Petten, & Besson, 1988). The N400 is sometimes followed by a brain potential that is more positive for words that are more difficult to integrate into text. These late positivities can vary in length and onset, but they usually begin around 500–700 ms following the onset of the word and can have a posterior or anterior distribution.

Research that has examined the N400 and Late Positivity components to the final words of literal and nonliteral statements have produced results that differ often from that found with reading and response time data (Coulson & Van Petten, 2002; Katz et al., 2004; Pynte et al., 1996). For example, Coulson and Van Petten (2002) examined the N400 and LPC to sentence final words when they completed a metaphor (*He knows that power is a strong intoxicant*), a literal sentence (*He knows that whiskey is a strong intoxicant*), and a literal mapping condition in which the final word was still used literally but involved more extensive conceptual mapping between the vehicle and topic (*He has used cough syrup as an intoxicant*). Their results showed that the N400 and LPC effects varied systematically as a function of how much abstraction was necessary for comprehension: sentence final words to metaphors produced the largest N400 and the largest LPC, whereas words that completed a literal sentence produced the smallest N400 and LPC, and words in the literal mapping condition fell between. These results have been taken to show that people have more difficulty integrating the figurative than literal meaning of words into sentence contexts and that the process of constructing literal and figurative meanings involve similar mechanisms. Coulson and Van Petten (2002) also interpreted their results as evidence that reading and response time measures are not as sensitive measures of figurative language processing as those obtained through ERP methodology.

Research by Ferretti et al. (2007b) provides additional support for Coulson and Van Petten's (2002) claims and extend previous ERP research by directly comparing self-paced reading times and ERPs for identical stimulus sets, and by examining slow cortical potentials in addition to N400 and LPC components. There are at least two advantages for examining slow potentials during figurative language processing. First, because the averages are time-locked to the first word and span the entire crucial statement, they provide a clear depiction of the differences between experimental conditions as they develop over the statements. When experimental effects emerge early in statements, time-locking averages to individual words that occur later in the statement make the interpretation of these single words problematic. Second, there is a growing body of evidence in the sentence processing literature that demonstrate slow potentials are sensitive to the ease in which sentences and clauses are integrated into mental representations of the text and, thus, have been taken as an index of working memory load (Ferretti, Kutas, & McRae, 2007a; Haarmann, Cameron, & Ruchkin, 2003; King & Kutas, 1995; Münte, Schiltz, & Kutas, 1998). These studies show that sentences and clauses that are more difficult to integrate produce more negative slow potential amplitudes than text that is easier to integrate and that these differences tend to be largest over anterior head locations. The advantages of slow potentials make them particularly useful for investigations in which the time-course and ease of figurative versus literal language interpretation are of primary interest (Ferretti et al., 2007b).

Ferretti et al. (2007b) investigated slow cortical potentials while people read familiar proverbs in discourse contexts biased toward either a literal or figurative interpretation. Recall the observation made earlier that people should find it easier to integrate familiar proverbs with literal contexts, because proverbs also tend to be valid literal statements (e.g., *you really shouldn't count your chickens before they hatch*). Thus, contexts biased toward the literal meaning are consistent with both the literal and figurative meaning of the proverbs. Ferretti et al. showed that at the third word of the proverbs, slow potentials at anterior regions of the head were more positive for proverbs preceded by literally biasing contexts compared with figurative biasing contexts, and this difference was sustained over the remaining words in the proverbs.

In a second experiment, Ferretti et al. (2007b) found, as did Katz and Ferretti (2001), no differences between the two contextual conditions in a word-by-word self-paced reading time study employing the exact same items used in the ERP study. The authors concluded that their research suggests people have less difficulty integrating familiar proverbs in literal than figurative contexts due to the greater amount of conceptual overlap in the literal contexts and, consistent with the claims of Coulson and Van Petten (2002), that self-paced reading is

not always as sensitive to text-integration differences as ERP methodology during figurative language processing. Ferretti et al. also concluded that the results cannot be accounted by the standard pragmatic approach, because context had an effect very early in the statements, or by the graded saliency model, because people had more difficulty integrating the salient figurative meaning than the less salient literal meaning into discourse contexts.

Explicit markers and proverb comprehension

The present study continues our investigation into how discourse constraints impact slow cortical potentials during figurative language comprehension by investigating how explicit markers, such as *literally speaking* and *figuratively speaking*, influence how people interpret proverbial statements. Explicit markers are brief statements presented immediately before a given statement that invite some pragmatic interpretation of that statement, specifically, “linguistically encoded clues which signal the speaker’s potential communicative intentions” (Fraser, 1996 p. 168). Fraser discussed several classes of pragmatic markers and, most importantly for our purposes, “commentary pragmatic markers” that serve the function of commenting on some aspect of the message. In the relevant subvariant of this class (which Fraser labels as “manner-of-speaking” markers), a speaker/writer informs the listener/reader about how the message should be understood. Thus, for instance, if a person were to introduce a message by the phrase “metaphorically,” for instance, the intent is to convey the message should not be taken as literally true. On face, such markers would be especially important in communicating one’s intent when using proverbs, because the interpretation of proverbs can be plausible in either a figurative or literal sense and, thus, the markers can help to disambiguate the intended meaning of the proverbial phrases.

To our knowledge, there has been only one study that has investigated the role that markers play in the interpretation of proverbs during online discourse comprehension. In this research, Katz and Ferretti (2003) examined the markers *Literally speaking*, *In a manner of speaking*, and *Proverbially speaking*, on self-paced reading times for familiar and unfamiliar proverbs placed in figurative or literal biasing contexts. These markers were employed, because they tend to be used in everyday language: a Google search of English web pages containing the aforementioned markers in text shows the phrase *In a manner of speaking* appeared the most (1.18 million), followed by *Figuratively speaking* (722,000), *Literally speaking* (274,000), and *Proverbially speaking* (25,800). Although these markers helped to disambiguate the meaning of unfamiliar proverbs, the influence of these markers on familiar proverbs was small. With familiar proverbs, the literal marker had no influence on reading times

for the literal-biased contextual condition, whereas the figurative markers increased reading times up to the second last word of the proverb. From the second-last word of the proverb through to the beginning of the subsequent sentence, people read the proverbs at a similar rate regardless of the contextual bias and regardless of type of marker employed. In the research reported here, we will examine again the role played by explicit markers in the processing of proverbs, this time employing ERP methodology. As noted above, the earlier study employed reading time methodology and found minimal effects of the markers on the processing of familiar proverbs. Given the importance of context effects on distinguishing between competing models of figurative language processing, we felt it appropriate to revisit the issue but now employing a more sensitive methodology. Moreover, recent work indicates that the terms “literally” and “literally speaking” may serve more than just a hint that the message should be taken at face value (Israel, 2002).

The present research contrasts a literal context that ends with the marker *literally speaking* with a figurative context that ends with the marker *figuratively speaking*. As noted above, Fraser (1996) indicates that a pragmatic marker, such as “figuratively speaking,” should signal to the listener/reader that the message should not be taken literally. Presumably, an analogous argument can be said of the use of the marker “literally”: the message should be taken as literal, and no other meaning need be inferred. Israel (2002) argues that, historically, the use of the marker “literally” was in fact used in that forewarning fashion but that more recent usage shows a shift from that function. Israel argues that the use of the marker “literally” is sensitive to contextual factors and can be used as somewhat of an intensifier and in other cases may not override a figurative interpretation, presumably as would be the case with salient nonliteral meanings. Barnden (2016) makes a case that “literally” often is used as an intensifier or as a way of marking the message as being conveyed in a hyperbolic manner. The one experimental investigation of the use of “literally” was by Givoni, Giora, and Bergerbest (2013) who, following the Graded Saliency Hypothesis described earlier, argued that there is a pragmatic need to mark the intended nonsalient use of a statement, because the salient meaning will be activated regardless of context. They argued further that a marker, such as “literally,” is a hint that the nonsalient meaning is intended and thus will serve only to have an effect on the nonsalient sense. In two off-line rating studies and an analysis of a corpus, they demonstrated that the presence of a marker invited people to decide that the intent was to convey the nonsalient sense (relative to a nonmarker control).

As extended to the current study, the effect of the marker “figuratively speaking” and “literally speaking” were examined with event-related potential methodology for both familiar and unfamiliar proverbs. The literature described above suggests that the effect of the marker “figuratively speaking”

should be quite straightforward as a hint that the intended meaning is the nonliteral meaning. The use of the “literally speaking” marker is more complex and may be taken as a means of informing the reader that the intended meaning is literal, and in some cases nonliteral (Israel, 2002), but is simultaneously imparting the information is being used in an exaggerated form (Barnden, 2016). However, if Givoni et al. (2013) is correct, the “literally” marker should only show an effect when the context indicates that the nonsalient use is being intended, especially for familiar proverbs used literally. Three studies are presented. In Experiment 1, we examined the influence of explicit markers on both familiar and unfamiliar proverbs presented in isolation. In this manner, we indexed the influence of the markers on proverb comprehension, independently of the rich discourse contexts used in the other experiments. In Experiment 2, we examined how the presence versus absence of the markers influenced comprehension of familiar proverbs when they were presented in figurative and literal contexts. In Experiment 3, we contrasted two different literal contextual conditions in addition to a figurative contextual condition. One literal condition had contexts without any content words that overlapped with the proverbial statements, whereas the second literal condition was identical with the exception that it contained content words that overlapped with the proverbial statements. There are a couple of reasons for contrasting these two literal conditions. One is that it enabled us to investigate how the overlapping content words may lead to specific expectations for the upcoming figurative statements. Furthermore, adding the overlapping condition provided a third level of possible integration difficulty due to the amount of abstraction that must occur for people to integrate conceptually the proverbial statements. That is, with content overlap, it may be easier to find the antecedent referent than when a synonym is employed.

Experiment 1

Experiment 1 examined the effect of markers on the processing of proverbs when there was not the added complexity of integrating these items into a larger discourse context. A second goal was to manipulate the salience of the figurative meaning of the proverbs by investigating the influence of the markers on both familiar and unfamiliar proverbs. The salient meaning of unfamiliar proverbial statements is the literal meaning, whereas for familiar proverbs the salient meaning is the figurative meaning (see Giora, 2003). If the marker provides useful information for comprehending the proverb, then we should find an interaction between the familiarity of the proverb and the type of marker employed. Specifically, the figurative marker should lead to more positive slow potentials for the familiar proverbs relative to when literal markers are used, because the figurative markers will be most consistent

with the salient meaning of those proverbs. It is unclear what will happen with the use of literal markers. If the literal markers merely provide information that a literal reading is intended (even in an exaggerated fashion), then the literal markers should lead to more positive slow potentials for unfamiliar proverbs, because these markers will be more consistent with the salient literal meaning of these proverbs. On the other hand, if the literal marker signals the nonsalient use is intended, then more positive slow potentials should be found with the familiar (and not unfamiliar) proverbs. Furthermore, the influence of the markers on slow potentials over the proverbs should be larger at anterior than at posterior head locations (Ferretti et al., 2007b).

Method

Participants

Twenty-eight undergraduate psychology students (17 females) from Wilfrid Laurier University participated for course credit. As in all experiments reported below, all participants were right-handed (as determined by the Edinburgh Handedness Inventory), neurologically normal, and were native English-speaking and had normal or corrected-to normal visual acuity.

Materials

Fifty familiar proverbs and 50 unfamiliar proverbs that were 7 words in length were selected from the familiarity rating study (1 = not at all familiar, 7 = very familiar) reported in Ferretti et al. (2007). The familiar proverbs ($M = 5.4$, range: 3.9–6.8) were rated significantly higher than the unfamiliar proverbs ($M = 2.3$, range: 1.4–3.8), $t(49) = 65.21$, $p < 0.001$. The proverbs and their literal (*Literally speaking*) and figurative markers (*Figuratively speaking*) were placed across two lists. Each list contained all proverbs with 25 of the items from each of the 4 experimental conditions (Familiar Proverb / Figurative Marker, Familiar Proverb / Literal Marker, Unfamiliar Proverb / Figurative Marker, Unfamiliar Proverb / Literal Marker). No participant saw any proverb or context more than once, and across the two lists each proverb was paired with a literal and figurative marker. We minimized expectation effects regarding the proverbs by including 100 sentences as part of each list, and none of these sentences contained figurative statements.

EEG Recording and Analysis The electroencephalogram (EEG) was recorded from 64 electrodes (including the 2 mastoid electrodes, a reference at the vertex, and a ground located between Fz and Fpz) distributed over the scalp according to the 10–20 placement standard. ECI electrolyte gel was used in conjunction with Ag/AgCl electrodes. EOG artifacts were

monitored via additional electrodes placed on the outer canthus of each eye and by electrodes placed below and above the left eye. Electrode impedances were kept below 5K Ω . EEG was processed through a Neuroscan Synamps2 amplifier and filtered with a bandpass of 0.05 Hz (6 dB/octave) to 100 Hz (6 dB/octave) and was digitized at 250 Hz.

The data were re-referenced offline to the average of the right and left mastoids. High-frequency noise was removed by applying a low-pass filter with a cutoff of 30 Hz (6 dB/octave). ERPs were then computed in epochs that extended from 200 ms before the first word of the sentence (i.e., literally or figuratively) to 500 ms after the final word's onset (–200 to 4,500 ms). The waveforms were baseline corrected before averaging.

Procedure

Participants sat in a chair in front of a computer monitor located in an electrically shielded chamber. They were instructed to read the words one at a time and to answer periodic comprehension questions (one third of trials had questions) by pressing buttons labeled “Yes” and “No.” The 100 experimental trials and 100 filler trials were presented one word at a time in the center of a computer screen. All words were presented for a duration of 300 ms with an SOA of 500 ms. The same procedure was used in Experiments 2 and 3.

Trials contaminated by blinks, eye-movements, and/or excessive muscle activity were rejected offline before averaging; a total of 28% of trials were lost due to such artifacts. Average waveforms were created for each condition at each electrode for every participant. Five-way ANOVAs were then conducted on the mean slow potential amplitudes at 9 time windows: one for each 500 ms word region in the proverbs, and one for each of the words comprising the markers. The primary factors of interest were context (figurative marker vs. literal marker), familiarity (familiar proverbs vs. unfamiliar proverbs), and anteriority (anterior vs. posterior electrode sites), all of which were within-participants variables. We only examined Anteriority as a topographical variable because our predicted differences in slow potentials between conditions was expected to be maximal at anterior head locations, as found by previous proverb research with similar stimuli (Ferretti et al. 2007b), and by previous reading research involving literal sentences (King & Kutas, 1995; Ferretti et al., 2007a). Furthermore, Ferretti et al.'s (2007b) topographical analysis showed that contextual variables only interacted with their anteriority factor. There were no interactions between contextual variables and their hemisphere and laterality factors. Their slow potential differences were also relatively similar at all anterior electrode sites, and no differences were found at any posterior sites. Therefore, to simplify the statistical analyses for the experiments reported below and to help

reduce type 1 error by restricting the total number of factors and comparisons between conditions (Luck & Gaspelin, 2017), we only included an anteriority factor with two levels (anterior vs. posterior).¹ As a result, the anteriority factor was created by dividing the electrode sites in half, with the electrode sites from the central to the prefrontal region of the head comprising the anterior condition, and the remaining electrodes over the back of the head comprising the posterior condition.²

As in all experiments reported herein, electrode was a within participant factor and list was used as a between participant factor to stabilize any variance caused by rotating participants across the different lists (Pollatsek & Well, 1995). Note that the list variable has no theoretical interest, so it is not discussed in the results reported below. All *p*-values in this and subsequent experiments are reported after Epsilon correction (Huynh-Feldt) for repeated measures with greater than two degrees of freedom. We also used a false discovery rate (FDR) with a corrected threshold set at $p = 0.05$ within each temporal region of interest. Table 1 displays the results of the ANOVAs for the primary factors of interest in each region of interest, and Figure 1 shows the mean amplitudes at anterior and posterior electrodes.³

Results

Analyses for Markers As shown in Table 1, the only effect that reached significance was a main effect of anteriority at the second word of the marker. This effect occurred because amplitudes for anterior electrode sites were more positive than for posterior electrode sites.

¹ We recognize that clustering electrodes into just anterior and posterior head locations is novel as many topographical analyses reported in the literature contrast different lateral electrode locations, and often involve more than 2 levels of anteriority. However, we feel that our analysis is appropriate as our experimental predictions are motivated by highly related research, and adding additional topographical variables that are not motivated would make our analysis less statistically sound.

² In order to have an equal number of electrode sites contributing to the anterior and posterior conditions (30 electrodes per condition), two lateral electrodes (T7, T8) in the central region were removed from the analyses in all experiments.

³ We used the G*Power3 program (Faul, Erdfelder, Lang, & Buchner, 2007) to compute post hoc power analyses for the crucial three-way interactions between familiarity, context, and anteriority in Experiment 1, and the crucial two-way interactions between context and anteriority in Experiments 2 and 3. The results of this analysis for Experiments 1, 2a, and 3 demonstrated power of ≥ 0.79 for all word regions. In Experiment 2b, power was > 0.90 for the two words that comprised the markers. For the word regions in the proverb, power ranged between 0.07 and 0.33. The low power found for the words in the proverb is primarily a result of the very small effect sizes (0.04 to 0.14; Cohen, 1977). For example, to achieve power of 0.80, 681 participants are needed for an effect size of 0.04, and 68 participants are needed for an effect size of 0.14. Although we cannot completely rule out the presence of two-way interactions between context and anteriority for the proverb regions in Experiment 2b, we can rule out that these interactions were comparable to those observed in Experiment 2a.

Table 1. Experiment 1 Anova results for each of the 9 word regions (500 ms Epochs) in the proverbs

Proverb region					
	1st Marker	2nd Marker	1st Word	2nd Word	3rd Word
F	F < 1	F < 1	F < 1	F < 1	F < 1
C	F < 1	F < 1	F < 1	F < 1	F < 1
A	F < 1	F(1,26) = 26.59**[.51]	F(1,26) = 70.65**[.73]	F(1,26) = 74.67**[.74]	F(1,26) = 64.28**[.71]
FxC	F < 1	F(1,26) = 2.84	F(1,26) = 3.56	F(1,26) = 2.92	F(1,26) = 2.60
FxA	F(1,26) = 2.64	F(1,26) = 3.20	F < 1	F < 1.40	F < 1
CxA	F < 1.67	F < 1	F < 1.33	F < 1	F < 1
FxCxA	F < 1.10	F < 1	F(1,26) = 3.53	F(1,26) = 3.48	F(1,26) = 4.75
		4th Word	5th Word	6th Word	7th Word
F		F < 1	F < 1	F < 1.21	F(1,26) = 3.69
C		F < 1	F < 1	F < 1	F < 1
A		F(1,26) = 63.11**[.71]	F(1,26) = 56.51**[.68]	F(1,26) = 47.47**[.65]	F(1,26) = 53.63**[.67]
FxC		F(1,26) = 6.65*[.20]	F(1,26) = 6.58*[.20]	F(1,26) = 10.21**[.28]	F(1,26) = 14.42**[.36]
FxA		F < 1.87	F(1,26) = 2.48	F(1,26) = 3.48	F(1,26) = 5.92*[.19]
CxA		F < 1	F < 1	F < 1	F < 1
FxCxA		F(1,26) = 7.08*[.21]	F(1,26) = 9.14*[.26]	F(1,26) = 10.47*[.29]	F(1,26) = 7.80*[.23]

Note that F = Familiarity, C = Context, and A = Anteriority. Effect size (partial eta-squared) for significant effects appear in square brackets

* $p \leq 0.05$; ** $p \leq 0.01$

Analyses for words in the proverbs As predicted, familiarity and context interacted and this interaction was significant starting at the fourth word of proverbs and was sustained through the remaining words of the proverbs. This interaction occurred because, for familiar proverbs, amplitudes at the last four words of the proverbs were either significantly or marginally more positive when preceded by the figurative marker than the literal marker, whereas for unfamiliar proverbs, amplitudes were significantly or marginally more positive when the proverbs were preceded by literal markers than the figurative markers. This interaction was modified by a three-way interaction between familiarity, context, and anteriority that was either significantly or marginally at the third word through the last word of the proverbs. This three-way interaction occurred because differences between the markers for both familiar and unfamiliar proverbs were larger at anterior electrode sites than at posterior electrode sites.

Anteriority also interacted with familiarity at the last word of the proverbs as the difference between amplitudes at anterior and posterior locations was larger for unfamiliar than familiar proverbs, although for both types of proverbs the difference was highly significant. Finally, amplitudes were more positive at anterior locations than at posterior locations for every word in the proverbs.

Discussion

The pattern of slow potential results at anterior electrode sites confirms the prediction that without discourse contexts it is

easier to integrate the salient meaning of the familiar proverbs with the figurative than literal marker, and that unfamiliar proverbs should be easiest to integrate when preceded by literal than figurative markers. Taken together, the results of Experiment 1 support the assumptions of the graded saliency model inasmuch as the salient meaning was integrated more easily with markers that were consistent with the salient meanings. They do not support the ancillary aspects of that theory, which makes the claim that the use of the literal marker should be facilitative for familiar (salient) proverbs.

Our present findings overlap with aspects of ERP research that has examined the interpretation of familiar French and Italian idioms (respectively, Laurent et al., 2006; Canal et al., 2017; Vespignani et al., 2010). In Laurent et al., the N400 was smaller at the final word of familiar idioms when those words were consistent versus inconsistent with the salient figurative meaning of the statement, a result that also supports the assumptions of the graded saliency model. Although our present research did not examine the N400 to individual words, previous research by Ferretti et al. (2007b) demonstrated overlap between the onset of proverb integration difficulty indexed by slow cortical potentials and the onset of differences in the N400. Our current research shows that in minimal discourse context (i.e., just the markers), the salient meanings of proverbial statements is easier to integrate than the less salient meaning and, importantly, this difference in integration difficulty occurred long before the final word of the proverb was reached.

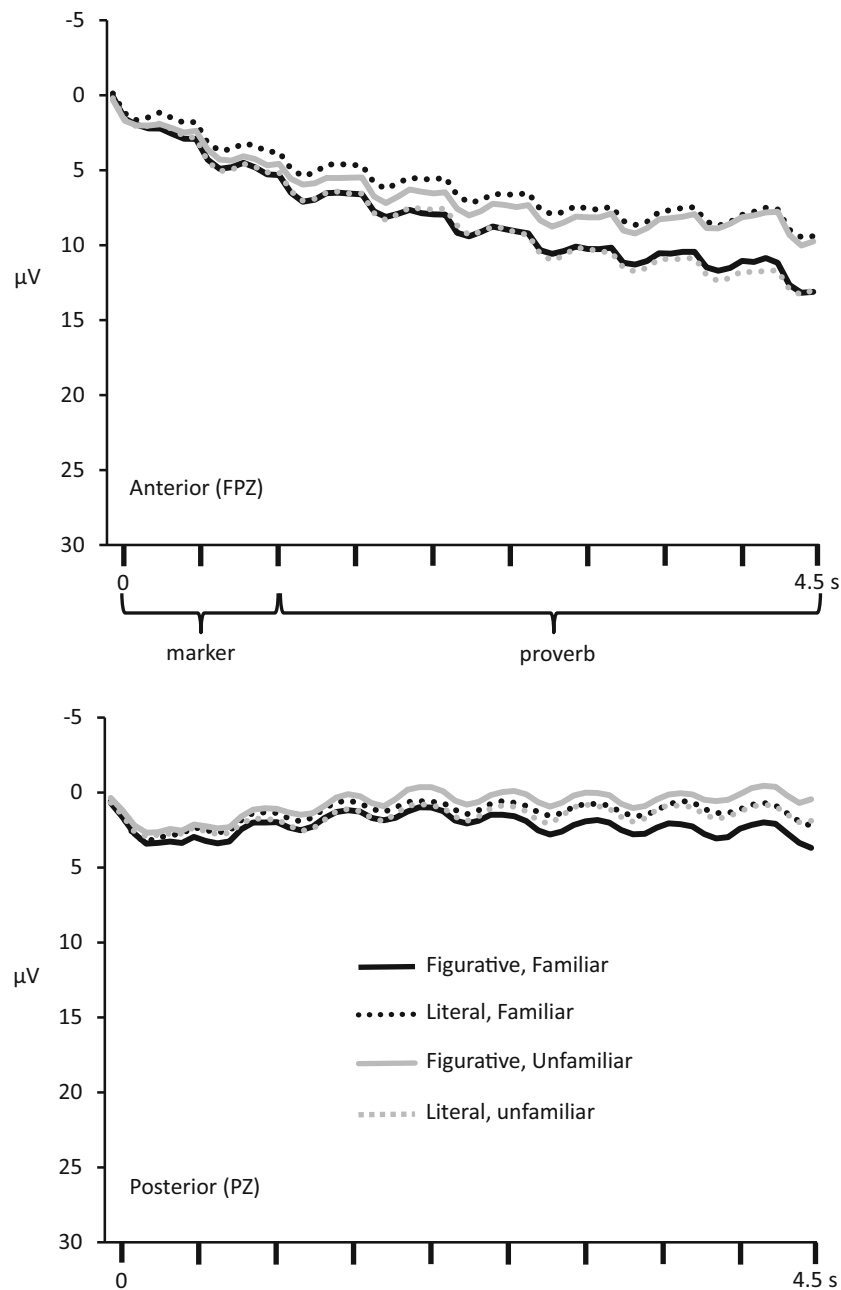


Fig. 1 Experiment 1 grand averages at anterior (FPZ) and posterior (PZ) electrodes. The amplitudes are shown after being filtered with a low pass filter set at 0.7 Hz to reveal the development of slow potentials over the markers and proverbs

Recently, researchers have argued for two processes at work at the level of fixed expression processing (e.g., idioms, proverbs, and other collocations): a probabilistic process (similar to constraint satisfaction) that is in play before the fixed expression is recognized, and a category-matching process in play after the insight occurs (Bianchi et al., 2019; Canal et al., 2017; Molinaro, Barraza, Carreiras, 2013; Molinaro & Carreiras, 2010; Vespignani et al., 2010). Given that our familiar proverbs are most likely to have a recognition point and our unfamiliar proverbs do not, one might expect from this perspective that

quite different effects might emerge for familiar versus unfamiliar proverbs, especially at the fourth word of the proverb where significant slow wave effects began to emerge. However, we found effects only consistent with a probabilistic mechanism in which expectations driven by a literal marker facilitated processing of the unfamiliar proverb (in which the literal sense is salient) and that expectations driven by a figurative marker facilitated processing of the familiar proverb (in which salient meaning is the proverbial sense). We discuss this issue further in the general discussion.

Experiment 2

In Experiment 2, we investigated the influence of the presence versus absence of explicit markers on slow potentials during the comprehension of familiar proverbs placed in figurative and literal biasing contexts. Experiment 2a examined the influence of the marker *figuratively speaking* on familiar proverbs presented in figurative contexts, and Experiment 2b examined the influence of the marker *literally speaking* on the same proverbs presented in literal contexts. Thus, in these conditions, the marker was always a valid cue for subsequent usage. Note that we ran two separate studies to maximize the number of passages per condition given the constraints that we had on the total number of available proverbs.

Based on previous self-paced reading results employing explicit markers (Katz & Ferretti, 2003) and recent ERP results that directly contrasted the literal and figurative context conditions without the aid of explicit markers (Ferretti et al., 2007b), we expect slow potentials at anterior head locations will show that people experience more difficulty integrating familiar proverbs in figuratively biasing contexts when the marker *Figuratively speaking* is present versus when it is absent. Presumably, this difficulty indicates effortful attempts to integrate the nonliteral sense. Recall that Ferretti et al. (2007b) demonstrated that slow potentials at anterior head locations were more positive for familiar proverbs presented in literal contexts than figurative contexts. The addition of the figuratively speaking marker could have at least two influences on proverb interpretation. First, the markers could cue the reader that an upcoming statement should be interpreted figuratively. In Ferretti et al. (2007b), differences between the figurative and literal conditions emerged by the third word. Thus, if the markers serve a cuing function, we might expect differences in amplitude between the two conditions to emerge earlier than the third word of the proverbs as a consequence of the marker. Second, based on Katz and Ferretti's (2003) reading time results for proverbs preceded by explicit markers, one could expect that the process of attempting to integrate the figurative meaning of the proverbs in figurative contexts is more difficult relative to a condition in which the markers are absent (Experiment 2a), but these differences might diminish by the second last word of the proverbs. In contrast, one position is that the literal marker will have a much smaller influence, if any, relative to when they are absent on slow potentials associated with familiar proverb processing; however, based on Givoni et al. (2013) we might find the literal marker might show larger positive slow potentials relative to the no-marker condition (Experiment 2b). Finally, in both experiments one might expect that people will have less difficulty interpreting the markers themselves, relative to comparison words that precede the figurative statement. Recall that the two-word phrases comprising the markers employed here are common English expressions, and that, in the present experiment,

participants encountered a marker on 15% (21/138) of the trials. If repetition and general familiarity of the markers contribute to the ease of reading the words, then the slow potentials may be more positive for the two words encountered just before the proverbs when they contain markers.

Method: Experiments 2a and 2b

Participants

Twenty-four undergraduate psychology students (14 females) participated for course credit in Experiment 2a, and a different set of 24 undergraduate psychology students (15 females) participated for course credit in Experiment 2b.

Materials and Procedure

The stimuli were constructed by pairing 42 familiar proverbs with either a figurative context (Experiment 2a) or a literal context (Experiment 2b). Each proverb was always preceded by four sentences that described conversations between people, and the sentence preceding each passage was identical across the different types of contexts unless a marker was present (see Examples 1a and 1b). In that case, the markers followed a pair of words that always described a person's name followed by the word *stated* (e.g., *Katherine stated "figuratively speaking..."* or *Katherine stated "literally speaking..."*). When the markers were absent, these two words always immediately preceded the proverbs. We ensured there was no lexical overlap between the content words that appeared in either the figurative and literal contexts with words in the target proverb. This ensured that any advantage found for the literal contexts could not be a result of lexical priming or other effects found with complete overlap between the words in the contexts and words in the proverbs. Because we intended in Experiment 3 to examine the role played by repeating words from the proverb in the discourse context, in the present experiment, we constructed an alternative literal version in which synonyms were replaced by words from the proverbs. Thus, for instance, with the proverb *the cat is out of the bag*, in the current study none of the content words were found in the discourse context for either the figurative or literal biasing condition. Thus, in the present study, we used a synonym for a word in the proverb (e.g., *sack* for *bag*). For the literal biasing contexts, there were 2.5 synonyms on average that overlapped with words in the proverbs.

(1a) Figurative context: "Why won't you tell me what you're making for my birthday?" said Joseph as he peered into the kitchen. "It's a surprise and you'll find out soon enough," said Katherine as she directed him away. As Katherine was pushing him from the kitchen entrance, a recipe for Beef Wellington, Joseph's favorite

food, dropped from the counter and he picked it up. “I guess I’ll find out sooner than you thought,” said Joseph. Katherine stated, “Figuratively speaking, the cat is out of the bag.”

(1b) Literal Context: “What could that possibly be?” wondered Joseph as he gazed at a strange-looking sack under the Christmas tree that appeared to be moving. “You’ll find out tomorrow,” said Katherine, as she moved to block his view of the sack. Suddenly, an animal scratched a large tear through a small air hole and climbed away. “I guess I’ll find out sooner than you thought,” said Joseph. Katherine stated, “Literally speaking, the cat is out of the bag.”

Three separate norming studies were conducted to ensure that our items were familiar (1 = not at all familiar, 7 = very familiar), were equal in how comprehensible they were in the literal and figurative contexts (1 = not at all easy to comprehend, 7 = very easy to comprehend), and differed in how figuratively or literally biasing the contexts were (1 = very literal, 7 = very figurative). The results of the normative ratings for the passages without markers were obtained from the norming studies reported in Ferretti et al. (2007b). To obtain the normative ratings when the markers were present, we had an additional 30 participants rate the items on the three dimensions described above. These norms also included ratings on each passage when the overlapping content words in discourse were present and when those words were replaced with synonyms. The ratings for the literal and figurative passages with and without the markers are presented in Table 2. The only statistical difference between the literal and figurative contextual conditions was for how figuratively or literally biasing the contexts were, and this was true when comparing items with markers ($t(41) = 19.13, p < 0.001$) and without markers ($t(41) = 15.39, p < 0.001$). Passages with markers were rated as easier to comprehend than the same passages without markers,

and this was true both for figurative contexts ($t(41) = 4.41, p < 0.001$) and literal contexts ($t(41) = 5.44, p < 0.001$). Finally, the two literal conditions with markers were rated equally familiar, comprehensible, and literal.

In Experiments 2a and 2b, the 42 proverbs and their corresponding discourse contexts were placed across 2 experimental lists. Each list contained all proverbs, 21 of the items contained markers and 21 items did not. No participant saw any proverb or context more than once, and across the two lists each proverb was presented with and without an explicit marker. Ninety-six passages that were similar in narrative form and length also were included as part of each list and none of these items contained figurative statements.

EEG Recording and Analysis The EEG recording and analysis parameters were the same as above, but note that in Experiments 2a and 2b, ERPs were time locked to the second last word that preceded the proverbs in the no marker condition. Thus, in both the marker and no marker condition the ERPs were always time locked to the second last word that preceded the proverbs.

Results (Experiment 2a, Figurative Contexts)

Trials contaminated by blinks, eye-movements, and excessive muscle activity were rejected offline before averaging; 28% of the trials were lost due to artifacts in Experiment 2a (27% were lost in Experiment 2b). Figure 2 shows the mean amplitudes at anterior and posterior electrodes. As can be seen, the mean amplitudes for the second word of the marker immediately preceding the proverb (i.e., *speaking*) were more positive than when the marker was absent. However, by the first word of the proverbial statement (i.e., 1,000–1,500 ms after the onset of the first word of the sentence) mean amplitudes for the no-marker condition become significantly more positive than the marker condition at anterior electrode sites, and this difference is maintained through to the end of the proverbial statement.

We conducted four-way ANOVAs on the mean amplitude for each condition at nine time windows of interest: one for each 500-ms word region in the proverbs, and one for each of the two words preceding the proverbs. The main factors of interest were context (presence of marker vs. absence of marker) and anteriority (anterior vs. posterior electrode sites), both of which were within participants variables. Table 3 shows the results for each region.

Analyses for marker region There was a marginal main effect of context at the second word and this occurred because amplitudes were more positive when the markers were present than absent. There also was a main effect of anteriority at the second word that occurred because amplitudes were more positive at anterior electrode sites than at posterior electrode sites.

Table 2. Mean ratings for the passages embedded in each of the 3 contextual conditions

	Figurative context	Literal context (synonym)	Literal context (overlap)
Without explicit marker			
Familiarity	5.0	5.0	5.0
Comprehension	5.5	5.4	N/A
Context figurativeness	5.7	2.2	N/A
With explicit marker			
Familiarity	5.0	5.0	5.0
Comprehension	6.1	6.2	6.1
Context figurativeness	5.9	2.1	2.0

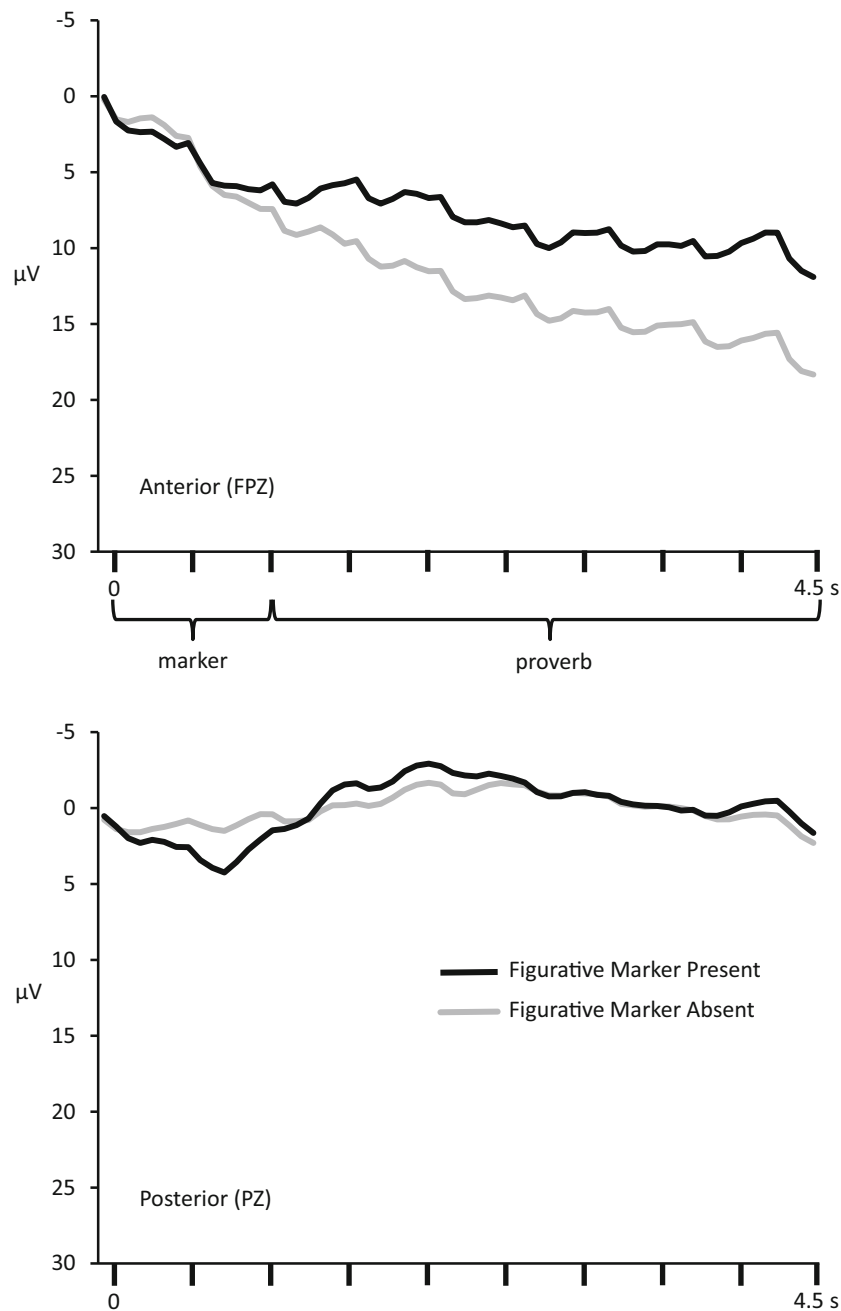


Fig. 2 Experiment 2a grand averages at anterior (FPZ) and posterior (PZ) electrodes. The amplitudes are shown after being filtered with a low pass filter set at 0.7 Hz to reveal the development of slow potentials over the markers and proverbs

Analyses for words in the proverb As illustrated in Table 3, the main effect of context was significant for the first three words and the seventh word of the proverbs. This effect occurred because mean amplitudes were more positive when the markers were absent than present. There also were significant context by anteriority interactions for all words of the proverbs. In each of these word regions, contexts without markers were more positive than contexts with markers across anterior electrode sites (all p 's < 0.001). Alternatively, there were no significant differences at

posterior electrode sites except for the second word of the proverbs ($p = 0.025$). Finally, there was a main effect of anteriority for every word region in the proverb. This effect occurred because amplitudes were more positive at anterior electrode sites than at posterior electrode sites.

Results (Experiment 2b, Literal Context)

Figure 3 shows the mean amplitudes at anterior and posterior electrodes. Table 4 shows the main effect of context,

Table 3. Experiment 2a Anova results for each of the 9 word regions (500 ms epochs) in the proverbial statements

	Context	Anteriority	Context x anteriority
<i>Word</i>			
2 nd word before	$F(1,22) = 2.98$	$F < 1$	$F(1,22) = 2.75$
1 st word before	$F(1,22) = 4.96$	$F(1,22) = 27.43^{**}[0.55]$	$F(1,22) = 3.10$
1 st	$F(1,22) = 4.65^{*}[0.17]$	$F(1,22) = 57.91^{**}[0.72]$	$F(1,22) = 6.81^{*}[0.24]$
2 nd	$F(1,22) = 18.08^{**}[0.45]$	$F(1,22) = 81.18^{**}[0.79]$	$F(1,22) = 19.49^{**}[0.47]$
3 rd	$F(1,22) = 7.50^{*}[0.25]$	$F(1,22) = 86.71^{**}[0.80]$	$F(1,22) = 25.13^{**}[0.53]$
4 th	$F(1,22) = 2.08$	$F(1,22) = 66.37^{**}[0.75]$	$F(1,22) = 26.47^{**}[0.54]$
5 th	$F(1,22) = 2.51$	$F(1,22) = 50.39^{**}[0.70]$	$F(1,22) = 18.58^{**}[0.46]$
6 th	$F(1,22) = 2.66$	$F(1,22) = 45.20^{**}[0.67]$	$F(1,22) = 22.14^{**}[0.50]$
7 th	$F(1,22) = 4.60^{*}[0.17]$	$F(1,22) = 47.05^{**}[0.68]$	$F(1,22) = 17.17^{**}[0.44]$

Effect size (partial eta-squared) for significant effects appear in square brackets. Note that in the marker condition, the last two words that preceded proverbs were composed of the markers, whereas in the no marker condition they were the final two words of the context

* $p \leq 0.05$; ** $p \leq 0.01$

anteriority, and the interaction between context and anteriority for each word region across the proverbial statement.

Analyses for marker region There was a significant main effect of context at the marker region and this occurred because amplitudes were more positive when the markers were present than absent. There also was a significant interaction between context and anteriority for both words. These interactions occurred because the difference between when the markers were present versus absent was larger over anterior than posterior electrode sites at the first word (anterior, $F(1,22) = 65.11$, $p < 0.001$; posterior, $F(1,22) = 16.40$, $p < 0.001$) and at the second word (anterior, $F(1,22) = 47.73$, $p < 0.001$; posterior: $F(1,22) = 16.08$, $p < 0.001$). Finally, there was a significant main effect of anteriority for both words, which occurred because amplitudes were more positive at anterior than posterior electrode sites.

Analyses for words in the proverbs There were no significant effects of context or context by anteriority interactions for any words in the proverbs. There was a significant main effect of anteriority at all words of the proverb. This effect occurred because amplitudes were more positive at anterior electrode sites than at posterior electrode sites.

Across these two studies, the analyses of the slow potentials demonstrated that the initial integration of the markers into the developing discourse contexts was easier than when the two words before the proverbs did not include a marker. This result supports the contention that there is an initial benefit to comprehension based on list-wise factors, such as frequency, familiarity, and repetition.

By the first word of the proverbial statements, slow cortical potentials at anterior electrode sites revealed that markers started to influence the ease in which people could integrate

the proverbs: when the contexts supported that proverbial sense of the proverbs (Experiment 2a), the marker *figuratively speaking* led to considerable integration difficulty that was sustained over the remaining words in the proverbs. Alternatively, for literal contexts (Experiment 2b), there was no significant difference in the ease with which proverbs were integrated, whether the markers were present or not. This pattern of results suggests that cueing people with a figurative marker leads to more difficulty in integrating the words in the familiar proverbs with preceding discourse, even though the cue is providing valid support for the intended and salient meaning of the proverb. The present ERP results are consistent with Katz and Ferretti (2003) who demonstrated that self-paced reading rates are longer following figurative markers relative to when the markers are absent, whereas the literal markers had little influence on reading times. The present results differ from the earlier study, however, because the difficulty integrating the proverbs into figurative contexts (versus literal contexts) was sustained from the first word throughout the remaining words in the proverbs.

The present findings build upon Experiment 1 by determining that people have more difficulty integrating the salient meaning of the proverbs with figurative contexts than literal contexts. That is, although a valid figurative marker by itself is facilitative on integrating the salient meaning (as shown in Experiment 1), access of the salient meaning by itself does not confer an advantage in integrating with the larger discourse context. In fact, integration is more difficult in this condition. As in Experiment 1, we find no evidence for the notion that the pragmatic marker “literally speaking” aids in the integration of the nonsalient meaning of the proverb, which in this case is when a familiar proverb is intended literally. We discuss these findings in more detail in the general discussion.

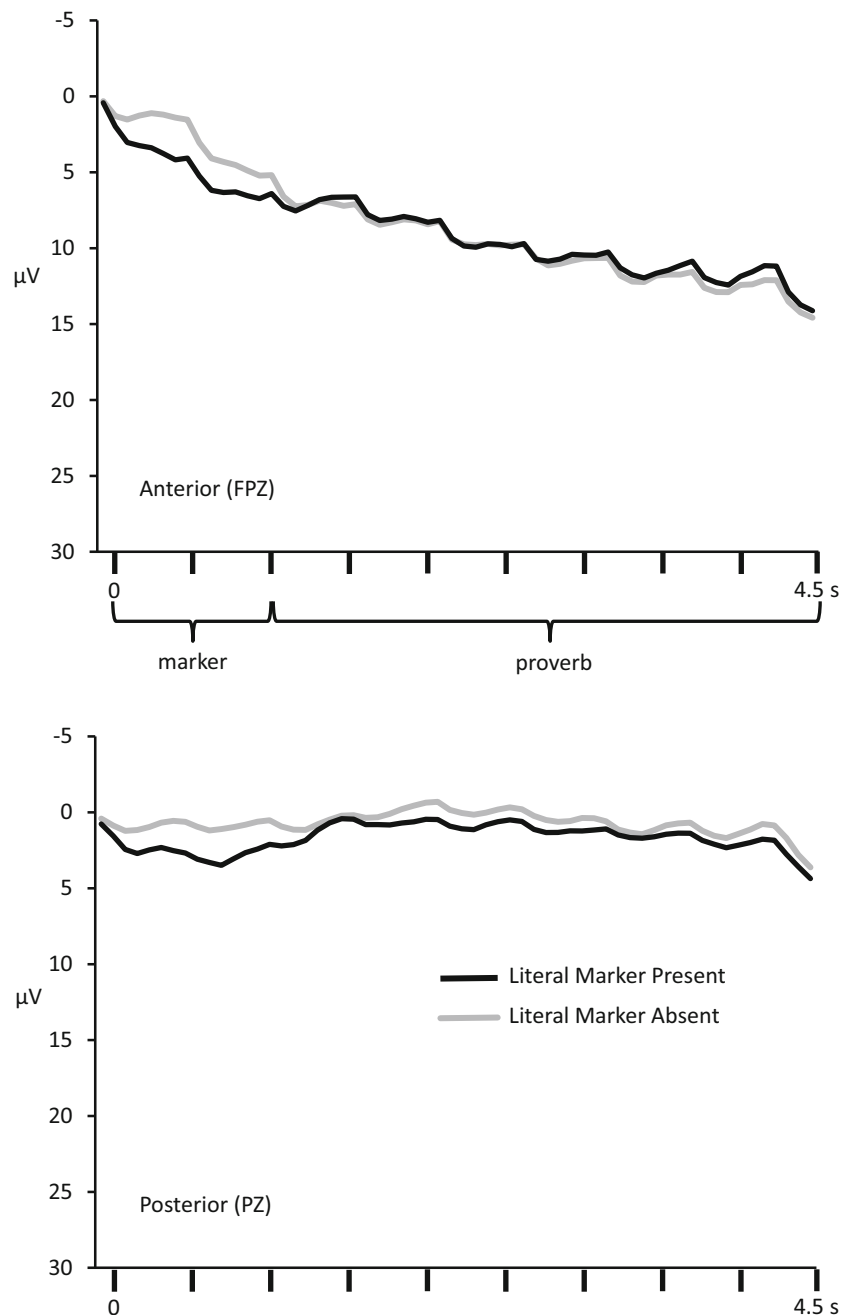


Fig. 3 Experiment 2b grand averages at anterior (FPZ) and posterior (PZ) electrodes. The amplitudes are shown after being filtered with a low pass filter set at 0.7 Hz to reveal the development of slow potentials over the markers and proverbs

Experiment 3

Experiment 3 extended the results from Experiment 1 and 2 in two important ways. First, we directly contrasted the influence of *Literally speaking* and *Figuratively speaking* on proverbs placed in literal and figurative contexts. Second, we included a third context condition that was identical to the literal condition used in Experiment 1 with the exception that the context contains content words that overlapped with the proverbial statements. Contrasting these two literal conditions (literal-synonym vs.

literal-overlap) enables us to investigate the role of lexical overlap on the ease of integrating figurative statements. If the use of proverb-specific words in contexts that support a proverbial continuation leads to expectations that a proverb is forth coming, then we may find differences between the two literal conditions that occur on the markers, before the proverbial statements are encountered. Moreover, if ease of integrating familiar proverbs into the discourse structure is in part based on the ease of finding antecedent referents and semantic links between words, we would expect to find a gradient in how easy it is to integrate

Table 4. Experiment 2b Anova results for each of the 9 word regions (500 ms epochs) in the proverb statements. Effect size (partial eta-squared) for significant effects appear in square brackets

	Context	Anteriority	Context x anteriority
<i>Word</i>			
2 nd word before	F(1,22) = 12.32**[0.35]	F(1,22) = 5.33*[0.19]	F(1,22) = 8.08*[0.27]
1 st word before	F(1,22) = 21.49**[0.49]	F(1,22) = 33.77**[0.64]	F(1,22) = 4.20[0.16]
1 st	F < 1	F(1,22) = 87.97**[0.80]	F < 1
2 nd	F < 1	F(1,22) = 90.73**[0.80]	F < 1
3 rd	F < 1.45	F(1,22) = 89.09**[0.80]	F < 1
4 th	F < 1	F(1,22) = 85.30**[0.79]	F < 1
5 th	F < 1	F(1,22) = 69.38**[0.75]	F < 1
6 th	F < 1	F(1,22) = 52.00**[0.70]	F < 1
7 th	F < 1	F(1,22) = 48.00**[0.68]	F < 1

Note that in the marker condition, the last two words that proceeded proverbs were composed of the markers, whereas in the no marker condition they were the final two words of the context

* $p \leq 0.05$; ** $p \leq 0.01$

the same statements into the developing discourse—easiest for the literal contexts with overlapping content words and hardest for the figuratively biasing contexts. These predictions are consistent with models that hold that figurative interpretations should be more difficult because of the difficulty in mapping between words and their preceding contexts (Coulson & Van Petten, 2002; Ferretti et al., 2007, b).

Method

Participants

Thirty-nine undergraduate psychology students (24 females) from Wilfrid Laurier University participated for course credit.

Materials and Procedure

The same 42 familiar proverbs and their figurative and literal contexts used in Experiment 2 were used in Experiment 3. In the present experiment, the additional literal context condition included content words that overlapped with the proverbs (see Example 2). Recall that on average, 2.5 content words overlapped with proverbial statements in this literal condition. The proverbial statements were preceded by the marker *literally speaking* for the two literal conditions, whereas the marker *figuratively speaking* always preceded the statements in the figurative condition.

(2) **Literal Contexts with Overlapping Words** (in bold): “What could that possibly be?” wondered Joseph as he gazed at a strange looking **bag** under the Christmas tree that appeared to be moving. “You’ll find out tomorrow,” said Katherine, as she moved to block his view of the bag. Suddenly, a **cat** scratched a large tear through a small air hole and climbed **out**. “I guess I’ll find out sooner than you

thought,” said Joseph. Katherine stated, “Literally speaking, the cat is out of the bag.”

The 42 proverbs and their corresponding passages were placed across three lists. Each list contained all proverbs with 14 of the items from each of the three experimental conditions. No participant saw any proverb or context more than once, and across the three lists each proverb was paired with each of the three types of context. The same 96 literal filler trials used in Experiment 2 were used in Experiment 3.

Trials contaminated by blinks, eye-movements and/or excessive muscle activity were rejected offline before averaging; a total of 22% of trials were lost due to such artifacts. Four-way ANOVAs were conducted on the mean amplitudes at nine regions of interest: one for each 500-ms word region in the proverbs, and one for each of the words comprising the markers. The primary factors of interest were context (figurative vs. literal-synonym vs. literal-overlap) and anteriority (anterior vs. posterior electrode sites), both of which were within-participants variables. Table 5 displays the main effect of context, anteriority, and the context by anteriority interaction for each region. Table 6 shows the results of the simple main effects for context at these regions. Figure 4 shows the mean amplitudes at anterior and posterior electrodes.

EEG Recording and Analysis The EEG recording and analysis parameters were identical to Experiments 1 and 2.

Results

Analyses for Markers As illustrated in Table 6, the analysis for the first word of the marker demonstrated that the mean

Table 6. Experiment 3 Anova results for each of the 9 word regions (500 ms epochs) in the proverbs

	Figurative vs. Literal-Overlap	Figurative vs. Literal-Synonym	Literal-Overlap vs. Literal-Synonym
<i>Word</i>			
1 st Marker	F(1,72) = 7.90*[0.10]	F < 1	F(1,72) = 5.36*[0.07]
2 nd Marker	F < 1	F(1,72) = 3.32	F(1,72) = 7.14*[0.08]
1 st	F(1,72) = 12.55**[0.15]	F < 1.75	F(1,72) = 4.93*[0.06]
2 nd	F(1,72) = 18.40**[0.20]	F(1,72) = 8.51**[0.09]	F(1,72) = 1.88
3 rd	F(1,72) = 12.09**[0.14]	F(1,72) = 3.81	F(1,72) = 2.33
4 th	F(1,72) = 11.70**[0.14]	F < 1.84	F(1,72) = 4.27*[0.06]
5 th	F(1,72) = 10.05**[0.12]	F < 1.13	F(1,72) = 4.45*[0.06]
6 th	F(1,72) = 11.00**[0.13]	F(1,72) = 2.36	F(1,72) = 3.17
7 th	F(1,72) = 10.89**[0.13]	F < 1.64	F(1,72) = 4.07

Effect size (partial eta-squared) for significant effects appear in square brackets

* $p \leq 0.05$; ** $p \leq 0.01$

amplitudes for the literal condition with overlapping words were more positive than the other two conditions, which did not differ from one another.

As illustrated in Table 5, there also was a significant context by anteriority interaction at the first word. This interaction occurred because the literal condition with overlap was more positive than the other two conditions at anterior (both p 's < 0.001) than posterior electrode sites (both p 's > 0.28). There was also a significant main effect of anteriority that occurred, because amplitudes were more positive over anterior than posterior electrode sites.

At the second word of the markers, the literal condition with overlap was significantly more positive than the literal condition with synonyms but was similar to the figurative condition. The figurative condition was marginally more positive than the literal synonym condition ($p < 0.08$). The interaction between context and anteriority was significant. This

interaction occurred because the literal condition with overlapping content words was more positive than the other conditions at anterior locations (both p 's < 0.01) but was similar at posterior locations (both p 's > 0.19).

Analyses for words in the proverbs Visual inspection of Figure 4 shows a gradient in how positive the slow potentials were across the seven words of the proverbs, especially at anterior electrode sites. Specifically, the amplitudes for the literal-overlapping condition were the most positive, followed by the literal condition without overlapping contents words, and the figurative condition was the least positive.

The main effect of context was significant for all words in the proverbs. At every word location, the literal condition with overlapping content words was significantly more positive than the figurative condition. The literal overlap condition also was either marginally significant or significantly more

Table 5. Experiment 3 Anova results for each of the 9 word regions (500 ms epochs) in the proverb statements

	Context	Anteriority	Context x anteriority
<i>Word</i>			
1 st Marker	F(2,72) = 4.50*[0.11]	F(1,36) = 26.33**[0.42]	F(2,72) = 7.07**[0.16]
2 nd Marker	F(2,72) = 3.73*[0.09]	F(1,36) = 76.74**[0.64]	F(2,72) = 6.08*[0.14]
1 st	F(2,72) = 6.41**[0.15]	F(1,36) = 26.33**[0.68]	F(2,72) = 6.59**[0.15]
2 nd	F(2,72) = 9.60**[0.21]	F(1,36) = 70.39**[0.66]	F(2,72) = 3.36*[0.09]
3 rd	F(2,72) = 6.07**[0.14]	F(1,36) = 74.44**[0.67]	F(2,72) = 3.35
4 th	F(2,72) = 5.93**[0.14]	F(1,36) = 70.65**[0.66]	F(2,72) = 4.81*[0.12]
5 th	F(2,72) = 5.21*[0.13]	F(1,36) = 63.79**[0.64]	F(2,72) = 6.44**[0.15]
6 th	F(2,72) = 5.51**[0.13]	F(1,36) = 57.19**[0.61]	F(2,72) = 6.52**[0.15]
7 th	F(2,72) = 5.54**[0.13]	F(1,36) = 65.29**[0.64]	F(2,72) = 5.85**[0.14]

Effect size (partial eta-squared) for significant effects appear in square brackets

* $p \leq 0.05$; ** $p \leq 0.01$

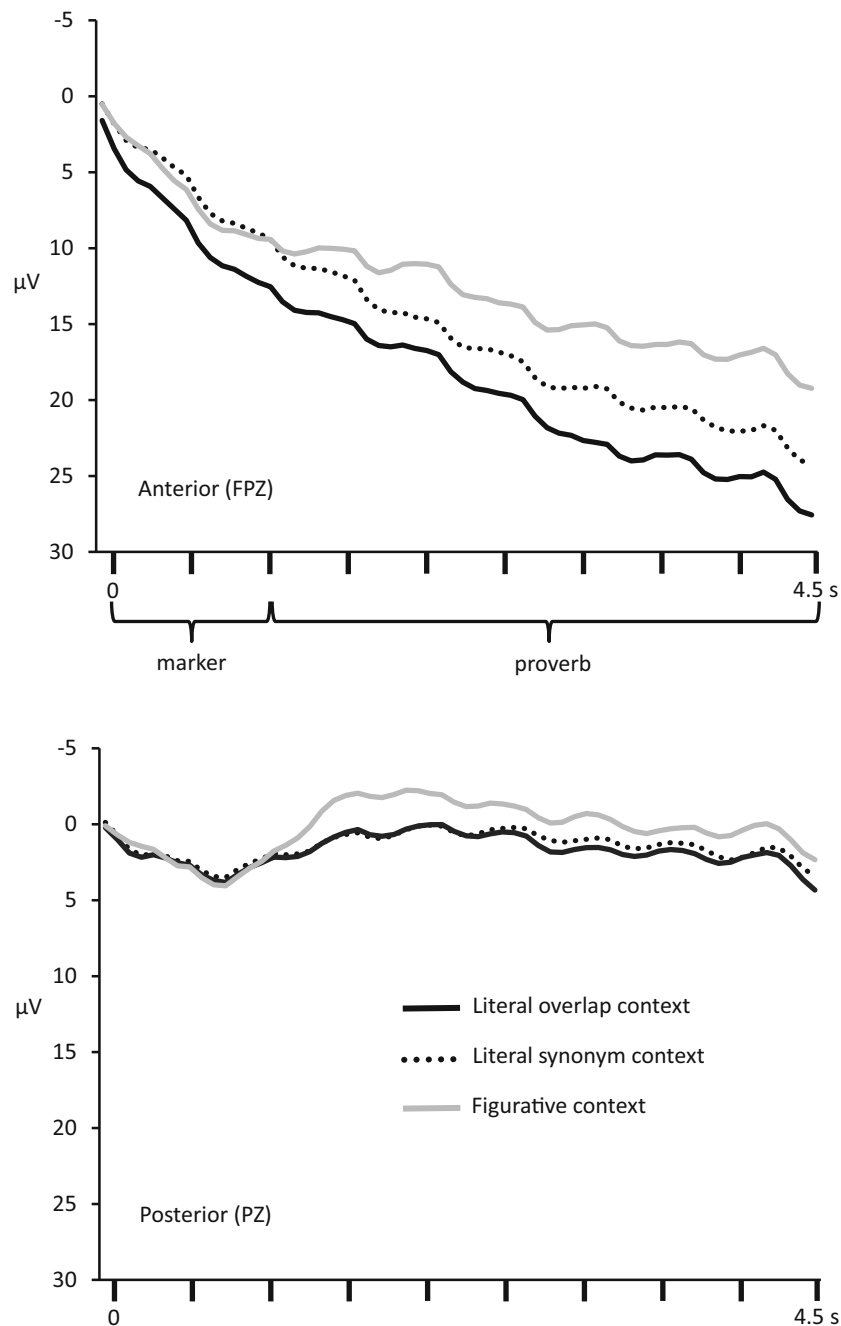


Fig. 4 Experiment 3 grand averages at anterior (FPZ) and posterior (PZ) electrodes. The amplitudes are shown after being filtered with a low pass filter set at 0.7 Hz to reveal the development of slow potentials over the markers and proverbs

positive than the literal synonym condition at the first word and then again from the fourth through seventh words. The literal synonym condition was significantly more positive than the figurative condition at the second word and marginally more positive at the third word.

Context and anteriority interacted significantly or marginally at all words of the proverbs. At all word locations, this interaction occurred, because the differences between the conditions were larger at anterior

than posterior electrode sites. Importantly, at anterior sites the literal overlap condition was significantly more positive than the other two conditions for all words of the proverbs (all p 's < 0.02), and the literal synonym condition also was significantly more positive than the figurative condition at all word regions except the first word (first word, $p > 0.19$; remaining 6 words, p 's < 0.03). Finally, at all word regions amplitudes at anterior electrode sites were more positive than at posterior sites.

Discussion

As predicted, slow potentials at anterior head locations were sensitive to the degree of conceptual overlap between the proverbs and the discourse contexts; the greater the lexical overlap between the words in the proverbs and the context the more positive the amplitudes. These results are consistent with ERP results that show that people have more difficulty integrating the figurative than literal meanings into contexts (Coulson & Van Petten, 2002; Ferretti et al., 2007b).

The literal condition with overlapping content words between the proverbs and contexts demonstrated that people used these few words to quickly generate expectancies for the proverbs. Specifically, the slow potentials for these markers were significantly more positive on the first word of the markers (i.e., *literally*) relative to the other two contextual conditions. This finding is striking when one considers that across the three contexts, the sentence preceding the markers were identical, and that the other literal condition only varied on average by 2.5 content words. Furthermore, the advantage of this condition over the other two conditions was found at the first word of the proverb, and relative to the figurative condition, remained throughout the proverbs. These results suggest that the key words (such as *cat* and *bag*) may have created an expectation about the nature of the sentence to come.

The results of Experiment 3 are most consistent with models that assume we actively construct interpretations during discourse processing, rather than retrieve entrenched meanings from semantic memory (Coulson & Van Petten, 2002; Katz & Ferretti, 2001). The fact that we find differences in slow potentials on the markers and for the first few words of the proverb cannot be accounted for by models of figurative language processing, which posit that the literal meaning of a statement must be processed before constructing a figurative meaning (Grice, 1975). The findings also are problematic for the graded salience hypothesis (Giora, 2003), which assumes obligatory access of salient meaning, which would need to explain why the familiar (and hence salient) proverbs were not as easy, and perhaps even easier, to integrate into the figurative context than into the literal contexts (especially when there were no overlapping content words). Our findings also are problematic for the ancillary hypothesis that the use of the literal marker should show an effect with familiar (i.e., salient) proverbs.

General Discussion

Our results demonstrated that explicit markers that are valid cues to an upcoming proverb facilitate the access of the meaning of that proverb when there is no discourse context. Thus, without any additional information it appears that people use the markers as hints on how to interpret the target statement. However, when there is discourse context—even one

consistent with the proverbial meaning—an explicit figurative marker confers no additional advantage to processing of the proverb, and paradoxically perhaps, actually makes the comprehension process more difficult. Similarly, the literal marker did not provide an additional advantage for processing the familiar proverbs when placed in literal contexts. On the other hand, unlike the figurative markers, there was no cost associated with the addition of the marker when they are embedded in contexts. Thus, our results demonstrate that one cannot study the influence of markers presented in isolation to determine how they will be used in a discourse context.

We have shown that one factor that makes the comprehension of proverbs in context difficult is due to a relative lack of easily accessible antecedents (such as found with lexical overlap) between the discourse and the proverb when used figuratively (compared to when it is used literally). This effect cannot explain the added difficulty created by use of the marker (compared with the exact same stimuli being presented without the marker in Experiment 2a). One possibility is that the use of markers in everyday usage might have pragmatic functions that lead the comprehender to expend more effort into seeking additional (or more nuanced) meaning from the proverb, especially one that is highly familiar. According to Givoni et al. (2013), the literal marker should highlight the less salient meaning of the familiar proverbs, and thus there should be a processing advantage with the addition of the literal marker. Our data are not consistent with this notion. We recognize the difficulty with interpreting this null effect, but one possibility is that highlighting the less salient meaning of the familiar proverb comes with a processing cost, thereby eliminating any advantage gained by employing the marker. We plan future experiments to address that possibility. Regardless, the interaction of discourse context with target sentences reinforces the argument that the processing of proverbs, or nonliteral language in general, provides insights into naturalistic language processing unaccounted by processing sentences without, or with minimal, context.

We have favored constraint satisfaction as a means of explaining the effects of the markers. This approach is not necessarily in conflict with an alternative explanation for the observed effects of the markers. One can frame the effects of the markers as setting up general expectations of what will follow in the discourse and thus be subject to the principles of communication proposed by Grice (1975), especially his maxim of quantity. The quantity maxim asserts that in attempting to be as informative as possible, one frames their communication to provide as much information as needed, not too much or too little. Thus, a marker is assumed to be informative and if, for example, appears to provide no added information to that available from the discourse context in general would invite the listener/reader to infer why the marker is being used at all. In principle, a Gricean approach would make predictions similar to what we observe here.

Despite the attractiveness of the Gricean perspective, there are several reasons that we view it as somewhat incomplete. In addition to the subtleties described by Israel (2002) regarding the nuanced nature of pragmatic markers, Grice has provided at best a general computational level description, namely identification of problems that people in communication must solve. He does not provide an algorithmic description of how the solution is implemented or physically represented. Indeed, there are some who argue that Grice's computational description is so general as to be nonimplementable (Frederking, 1996). To our knowledge, even after 35 years since Grice's seminal paper, there is scant evidence for successful implementation of his maxims. In contrast, constraint satisfaction is an algorithmic approach used to solve problems in resolving ambiguity, including linguistic ambiguities, makes clear predictions and has been instantiated in several domains. On that basis we are more comfortable with a constraint satisfaction approach, although we recognize that as Grice's maxims become more specified, constraint satisfaction algorithms, might be an approach eventually to their implementation.

Previous ERP research on figurative language processing has primarily involved examining the N400 and LPC components to sentence final words that are consistent with either a literal or figurative meaning of those sentences. This research is important, because it shows that people have more difficulty integrating the figurative meaning of a word into those sentence contexts, unlike many reading time and response time research that often finds no differences in the ease and timing of figurative and literal language processing. Of course, in real-world situations it is relatively rare that figurative statements are presented with such minimal contexts, and thus, it is important to examine figurative statements in larger discourse contexts, such as those employed in the present research.

To our knowledge, only the present research and the recent research by Ferretti et al. (2007b) have examined the influence of discourse contexts on slow cortical potentials. Although less is known about these brain potentials relative to other more common ERP components, previous research examining sentence processing has indicated that these brain potentials at anterior head locations are sensitive to integration difficulty for multiple words and are highly correlated with working memory capacity (Ferretti et al., 2007, a; King & Kutas, 1995; Münte et al., 1998). The present work, along with the research conducted by Ferretti et al. (2007b), shows that these potentials are sensitive to differences in the ease of integrating familiar figurative statements into discourse contexts, even when no differences tend to be found in self-paced reading. Furthermore, this research shows that examining slow potentials is useful for investigating the time-course of discourse constraints on the interpretation of sentences, particularly when differences between conditions arise early in sentences and influence processing of subsequent words. In our case, examining these potentials has provided insight into models of how and when people construct figurative interpretations online.

Fixed expressions have received more attention recently because of results that suggest contextual expectations consistent with a probabilistic process occur prior to the recognition of the expressions, followed by a category matching processes that occurs following recognition (Bianchi et al., 2019; Canal et al., 2017; Molinaro, Barraza, Carreiras, 2013; Molinaro & Carreiras, 2010; Vespignani et al., 2010). In research on idioms, a categorical match between a word with the preactivated string of lexical items elicits a posterior P300, and this electrophysiological response has been taken to reflect low-level perceptual processes during reading (see Kok, 2001, for a detailed discussion of the cognitive and electrophysiological basis of this matching process). Note that the P300 partially overlaps with the N400 in latency and topographic distribution, and this can lead to an extremely reduced N400 for the "categorically matching" word, particularly at posterior head locations. Bianchi et al. (2019) also have found N400 differences between high and low predictable words in sentences without fixed expressions that were not found to words that followed recognition points in proverbs.

The goal of our research was to examine how discourse constraints impact slow wave potentials during figurative language comprehension. As such, we examined the development of slow wave potentials over the proverbs and not the more commonly examined N400 and positive components that are elicited to individual words. However, our slow wave results in Experiment 1, which contrasted familiar proverbs (have recognition points) with unfamiliar proverbs (have no recognition points), showed that the contextual expectations created by the markers led to processing of the proverbs in a manner more consistent with a probabilistic than a categorical matching process. Sustained slow waves also are known to overlap in latency with the P300, but these two components can be differentiated based on their distribution over the scalp, with the P300 component having a more posterior distribution (Kok, 2001). The slow wave potentials in our present and past proverb research (Ferretti et al., 2007b) show a clear anterior distribution. Our results suggest that the slow wave potentials in the present study do not reflect contextual expectations that lead to a categorical matching process. However, more research is clearly needed and planned to clarify the relationship between slow wave potentials and the different types of contextual expectations that are created during the processing of sentences that contain fixed expressions.

Conclusions

Our results are problematic to both the standard pragmatic and graded salience models of nonliteral language processing. We demonstrated that people construct figurative meanings very early in proverbial statements and that the most salient meaning of a proverbial statement

(i.e., figurative meaning) is more difficult to interpret in discourse contexts than the less salient meaning (literal meaning). Our results are inconsistent with the claim that arises from graded salience theory that the marker should benefit the less salient meaning for familiar proverbs, at least with the online measure used here. It remains possible that offline consideration of the stimuli might replicate their findings and, if so, would indicate when people examine markers in leisure they might come to different understandings than when they are initially reading the material, as done in all three studies we report. Finally, our findings show that one difficulty in constructing context-appropriate figurative interpretations (relative to constructing context-appropriate literal interpretations) is based on lesser conceptual overlap present in figuratively biasing than in literally biasing contexts (Coulson & Van Petten, 2002; Ferretti et al., 2007b).

Author Notes This research was supported by a Canadian Foundation for Innovation (CFI) grant held by the first author and by separate NSERC Discovery Grants held by the first and second authors.

Open Practice Statement None of the data or materials for the experiments reported here are available, and none of the experiments was preregistered.

References

- Barnden, J. (2016). A Hyperbole-Based Account of the Intensificatory Usage of “Literally”. United Kingdom Cognitive Linguistic Conference UKCLC. University of Bangor, July 2016
- Bianchi, B., Shalom, D. E., & Kamienkowski, J. E. (2019). Predicting Known Sentences: Neural Basis of Proverb Reading Using Non-parametric Statistical Testing and Mixed-Effects Models. *Frontiers in Human Neuroscience*, 13:82.
- Canal, P., Pesciarelli, F., Vespignani, F., Molinaro, N., & Cacciari, C. (2017). Composition and Enriched Integration in Idiom Processing: An EEG study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 928-943.
- Blasko, D. G. (1999). Only the tip of the iceberg: Who understands what about metaphor? *Journal of Pragmatics*, 31, 1675-1683.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev.). New York: Academic Press.
- Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory and Cognition*, 30, 958-968.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Ferretti, T. R., Kutas, M., & McRae, K. (2007a). Verb Aspect and the Activation of Event Knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 182-196.
- Ferretti, T. R., Schwint, C. A., & Katz, A. N. (2007b). Electrophysiological and Behavioral Measures of the Influence of Literal and Figurative Contextual Constraints on Proverb Comprehension. *Brain and Language*, 101, 38-49.
- Fraser, B. (1996). Pragmatic markers. *Pragmatics*, 6, 167-190.
- Frederking, R. E. (1996) Grice's Maxims: “Do the Right Thing”. Presented at the Computational Implicature workshop at the AAAI-96 Spring Symposium Series, Stanford, 1996.
- Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language*, 37, 331-355.
- Gibbs, R. W. (1994). *The poetics of mind*. Cambridge, England: Cambridge University Press.
- Gibbs, R. W., Bogdanovich, J. M., Sykes, J. R., & Barr, D. J. (1997). Metaphor in idiom comprehension. *Journal of Memory and Language*, 37, 141-154.
- Giora, R. (2003). *On our mind: Salience, context and figurative language*. Oxford, England: Oxford University Press.
- Givoni, S., Giora, R., & Bergerbest, D. (2013). How speakers alert addressees to multiple meanings. *Journal of Pragmatics*, 48(1), 29-40.
- Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (eds) *Syntax and Semantics*. Vol. 3: *Speech Acts* (pp. 41-58). New York: Academic Press.
- Haarmann, H. J., Cameron, K. A., & Ruchkin, D. S. (2003). Short-term semantic retention during on-line sentence comprehension: Brain Potential evidence from filler-gap constructions. *Cognitive Brain Research*, 15, 178-190.
- Israel, M. (2002). Literally speaking. *Journal of Pragmatics*, 34(4), 423-432.
- Katz, A. (1966). Experimental psycholinguistics and figurative language. Circa 1965. *Metaphor and Symbolic Activity*, 11, 17-37.
- Katz, A. N., Blasko, D. G., & Kazmerski, V. A. (2004). Saying what you don't mean: Social influences on sarcastic language processing. *Current Directions in Psychological Science*, 13, 186-189.
- Katz, A. N., & Ferretti, T. R. (2001). Moment-by-moment comprehension of proverbs in discourse. *Metaphor and Symbol*, 16, 193-221.
- Katz, A. N., & Ferretti, T. R. (2003). Reading Proverbs in Context: The Role of Explicit Markers. *Discourse Processes*, 36, 19-46.
- King, J., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7(3), 376-395.
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, 38, 557-577.
- Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, 207, 203-205.
- Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology*, 69, 218-233.
- Laurent, J., Denhières, G., Passerieux, C., Iakimova, G., & Hardy-Baylé, M. (2006). On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research*, 1068, 151-160.
- Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, 54, 146-157.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283-312.
- Molinaro, N., Barraza, P., & Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *Neuroimage*, 72, 120-132.
- Molinaro, N., & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration

- during the processing of collocations. *Biological Psychology*, 83, 176-190.
- Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order: an electrophysiological analysis. *Nature*, 395, 71-73.
- Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 785–794.
- Pynte, J., Besson, M., Robichon, F., & Poli, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain and Language*, 55, 293-316.
- Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.
- Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive Mechanisms in Idiom Comprehension. *Journal of Cognitive Neuroscience*, 22, 1682-1700.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.