Published in final edited form as:

Neuropsychologia. 2014 November ; 64: 41–53. doi:10.1016/j.neuropsychologia.2014.09.012.

Word-to-text integration: Message level and lexical level influences in ERPs

Joseph Z. Stafura^{1,2} and Charles A. Perfetti^{1,2}

¹Department of Psychology, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260

²Center for the Neural Basis of Cognition, Pittsburgh, PA

Abstract

Although the reading of connected text proceeds in a largely incremental fashion, the relative degree to which message level and lexical level factors contribute to integration processes across sentences remains an open question. We examined the influence of both factors on single words using event-related potentials (ERPs). Word pairs with either strong or weak forward association strength were critical items: embedded as coreferential words within two-sentence passages in a text comprehension task, and as isolated word pairs in a word meaning judgment task. While the N400 ERP component reflected an effect of forward association strength on lexico-semantic processing in the word task (i.e., reduced N400 amplitudes were seen for strongly associated pairs relative to weakly associated pairs), in the comprehension task, passages embedded with any associated word pairs elicited reduced N400 amplitudes relative to coherent baseline passages lacking one of the critical words. These comprehension effects reflect responses from the highest skilled comprehenders. The results demonstrate the effects of message level factors, and reading abilities, on the processing of single words.

Keywords

text comprehension; integration; lexical association; word meaning; ERPs

1.0 Introduction

Text comprehension processes operate on a single word to connect its meaning with the reader's understanding of the text. These word-to-text integration processes are essential to building and updating, within and across sentence boundaries, a situation model of the text (Kintsch, 1988). These integrative processes are observable in word reading times (Tyler & Marslen-Wilson, 1977), eye-movements (Rayner, Sereno, Morris, Schmauder, & Clifton,

Correspondence concerning this article should be sent to either author at the Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260. jzs48@pitt.edu; Perfetti@pitt.edu, Phone: 412-624-7055; Fax: 412-624-9419.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1989), and evoked brain potentials that reflect lexico-semantics (Kutas & Hillyard, 1980). Thus, a tight coupling of the message level (the meaning of the text) and the lexical level (the meaning of the word) produces fluid word-to-text integration.

An important question is how this tight coupling comes about. In particular, how do the lexical level and the message level interact during text comprehension? To what extent do word-to-word connections in associative or semantic memory drive certain integrative processes? To what extent is the message level, which selects the text-relevant meaning, in control of the process? The current experiment examines a specific form of this general question: Does associative strength between coreferential words function *across sentence boundaries* to influence on-line word-to-text integration? If so, how does this influence compare to the effects of associative strength on word-to-word processing? Finally, does reading ability modulate effects of associative strength in word-to-text and word-to-word processing in similar ways?

The analysis of Event-Related Potentials (ERPs) measured on specific words provides a powerful method for investigating questions in text processing. The fine temporal correlation between the EEG signal and mass neuronal activity affords a millisecond by millisecond record of processing that is unavailable to other non-invasive measures. Much of ERP research has focused on the influence of context on the processing of words in isolated sentences, and to a lesser extent, in connected text. The combination of ERP measures and careful experimentation has been used to build theoretical models of language processing (e.g., Federmeier, 2007).

One particularly well-documented ERP measure in the study of language is the N400 component (Kutas & Hillyard, 1980), a negative-going deflection of the ERP waveform, peaking at around 400 ms after the onset of any potentially meaningful stimulus (Kutas & Federmeier, 2011). The initial discovery of the N400 revealed that it is larger (i.e., of greater amplitude) in response to words that are incongruent within their context relative to those that are congruent within their context (Kutas & Hillyard, 1980). Subsequently, the sensitivity of the N400 component to a large range of linguistic manipulations has been tested, including the cloze probability of words (Kutas & Hillyard, 1984) and their position in sentences (Kutas, Van Petten, & Besson, 1988; Van Petten & Kutas, 1990). At the lexical level, decreased N400 responses were found to words following lexical associates in sentences (Van Petten, 1993) as well as to words in the same conceptual category as expected words (Federmeier & Kutas, 1999).

ERP studies of word-by-word processing also have been carried out with connected texts. For example, van Berkum, Hagoort, & Brown, (1999) observed ERPs while participants read either single sentences or short texts in which the 3rd sentence was either congruent or incongruent. The critical words in the texts were congruent with the local sentence context, but incongruent with the message level context set up by the first two sentences. In both single sentences and three-sentence texts, the N400 on incongruent words was larger than on congruent words. The N400 effects (i.e., Incongruent - Congruent) and topographies were largely consistent across sentence and text reading conditions. These results demonstrate that

discourse-level meaning can influence the semantic processing of individual words in a manner similar to sentence-level meaning.

Other studies suggest discourse context allows the prediction of individual lexical items (van Berkum, Brown, Zwisterlood, Kooijman & Hagoort, 2005; Wicha, Morena, & Kutas, 2004). van Berkum et al (2005) studied native Dutch speakers who either listened to (Experiment 1) or read (Experiment 3) two-sentence passages, in which the first sentence was highly constraining for a specific noun in the second sentence. The experimental manipulations were the inclusion of the expected (congruent) noun or an unexpected (incongruent) noun in its place, as well as the inclusion of a consistently or inconsistently gender-marked preceding adjective. For example, the Dutch noun for “painting” (*schilderij*) has a neuter gender and could be preceded either by a consistent neuter gender “zero” suffix adjective (“big” = *groot*), or an inconsistent common gender –e suffix adjective (*grote*). As expected, the congruent, expected nouns elicited reduced N400 amplitudes relative to unexpected, incongruent nouns. More interesting, there was an effect on the preceding adjective. If the gender marking of the adjective was not consistent with the gender of the expected noun, an effect was observed on an early positive deflection between 50–250ms after adjective inflection onset. This, along with other evidence (Lau, Almeida, Hines, & Poeppel, 2009; Wicha et al., 2004), seems to indicate that the message level context can lead to the anticipation of specific words, and not simply abstract meaning features (e.g., Federmeier & Kutas, 1999).

To this point, most research on context effects have used at least moderately constraining texts, and has examined differences between processing contextually congruent and contextually incongruent words. To extend our understanding of the way multiple levels of representation interact, it is critical to examine message level and lexical level factors in the processing of specific discourse devices that connect words and texts in the construction of situation models. Anderson and Holcomb (2005) provide an example in a study (experiment 2) in which participants read two-sentence texts: the first sentence contained a noun in the object position that was repeated or synonymous with the word in the subject position in the second sentence. This second sentence subject was made either coreferential by the definite article (“the”) or new to the discourse by the indefinite article (“a”). The ERP measures on the critical words revealed that repetitions and synonyms elicited reduced N400 responses relative to filler words, with synonyms eliciting a N400 response between that of repetitions and fillers. However, the authors did not find a reliable N400 effect of coreference (“a” vs “the”) on the critical noun, which they took to suggest that the repetition and synonym effects were lexical in nature.

In an ERP study on word-to-text integration processes, Yang, Perfetti, and Schmalhofer (2007) demonstrated an effect on word processing driven by the referential availability of a critical word across a sentence boundary. For example, in their *explicit* condition the critical word was a repetition of a word in the first sentence (with occasional morphological variation; e.g., *exploded* - *explosion*), and in the *paraphrase* condition the critical word was conceptually related to an event expressed by a different word or phrase in the first sentence (e.g., *blew up* – *exploded*). Importantly, in contrast to Anderson & Holcomb (2005), the coreferential paraphrase words were not chosen to be synonyms of the antecedent words.

During reading of the critical words, ERP measures revealed reduced N400 responses for both repetition and paraphrase conditions relative to a baseline condition. A condition that did not contain a readily available antecedent in the first sentence, but required additional inferencing, did not elicit the same N400 reduction.

We can specify the processes involved in connecting the two sentences word-by-word by referring to the reading of the key word *explosion* from the Yang et al. (2007) study. In the *explicit* condition, the reader has constructed a situation model (Johnson-Laird, 1980; Kintsch, 1988) that includes a bomb explosion event from the final clause of the first sentence (“...the bomb hit the ground and exploded”). Integration of the word *explosion* is well supported both by the explosion event and the word *exploded* in the previous sentence. In the *paraphrase* condition, however, only the event structure (the event described by “blew up” in the first sentence) is available for integration—there is no word form overlap. As the reader encounters the “explosion” in the paraphrase condition, integration depends on making a coreferential link to the event described by “blew up” in the first sentence. It is this coreferential process that is captured by the phrase “word-to-text integration” and is responsible for a reduced N400 in the paraphrase condition. In the *baseline* condition, there is no ‘explosion’ event in the first sentence and thus no coreferential integrative process in the second sentence at the word “explosion”. Instead, the reader may establish a new referent (the explosion). However, even here, the word “bomb” appeared in the first sentence, which allows a word-level connection to be made when “explosion” is read in the second sentence. Thus, the advantage of the paraphrase condition (its N400 reduction) over the baseline condition is not dependent on the word “bomb” but seems to require a message level explanation in the form of referential binding.

While existing research using brief discourse contexts has provided evidence of message level factors on word processing, it has examined the lexical-level factors that might be involved in word-to-text integration to a lesser extent. One lexical level factor concerns the connections among words stored in memory, a factor that can be indexed by traditional associative norms that provide estimates of strength of association between two words or by directionless metrics that capture multi-dimensional semantic distances between pairs of words measured from large corpora, e.g. LSA (Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). Words preceded by semantically or associatively related words are processed more quickly and accurately than words preceded by unrelated words (Balota & Lorch, 1986; Meyer & Schvaneveldt, 1971). Such effects are thought to result from automatic spreading activation at short stimulus-onset asynchronies (SOAs) and the development of semantic expectancy sets at longer SOAs (Neely, 1991; Neely & Keefe, 1989). In certain contexts, associative priming is graded; Coney (2002) found a linear decrease in lexical decision reaction times with increasing associative strength between primes and targets. Priming effects have been found in ERP studies, where words preceded by related words elicit reduced N400s compared to words preceded by unrelated words (Bentin, McCarthy, & Wood, 1985; Holcomb, 1988).

Findings of lexical level influences during on-line text processing are somewhat mixed (Ledoux, Camblin, Swaab, & Gordon, 2006). On the one hand, lexically associated word pairs within sentences (isolated and in longer texts) have been found to elicit reductions of

the N400 component (Anderson & Holcomb, 2005; Baggio, van Lambalgen, & Hagoort, 2008; Camblin, Gordon, & Swaab, 2007; Van Petten, 1993). Additionally, lexical association effects have been found in eye-tracking measures, including reduced re-reading times and increased skipping rates on words following associatively related antecedents (Camblin et al., 2007), as well as reduced fixation times on those words (Carroll & Slowiaczek, 1986). On the other hand, several eye-tracking studies examining the effect of schematically-related terms in sentences have failed to find effect (Morris, 1994; Traxler, Foss, Seely, Kaup, & Morris, 2000). Some ERP studies also have not found the effects of lexical association for pairs of words embedded in sentences (Otten & van Berkum, 2008; Van Petten, Coulson, Weckerly, Federmeier, Folstein, & Kutas, 1999). In general, the effects of lexical level factors such as word association are greater in simpler materials or in those lacking congruence (Coulson, Federmeier, Van Petten, & Kutas, 2005). This attenuation of lexical level effects in rich contexts is consistent with the view that the comprehension of text involves abstract meaning coding through propositions abstracted from specific lexical instantiations (Kintsch, 1988).

However, with the exception of Anderson and Holcomb (2005), on-line studies of the relative importance of lexical and message-level factors have relied on associated words *within the same sentence*, words that, to a large extent, are not coreferential in the sense of the materials used in studies such as Yang et al. (2007). Although the message level has been manipulated through global discourse coherence, the lexical level has been examined through associated words within a sentence, and often part of the same clause (e.g., 1–3 words apart in Camblin et al., 2007). These conditions are useful for examining automatic spreading activation accounts of associative effects in context (Gouldthorp & Coney, 2011), but they do not address the cross-sentence integration issue we raise here. The paradigm developed by Yang et al. (2005; 2007) provides a useful test of readers' reliance on lexical level factors in word-to-text integration. Word association strength can be varied across sentence boundaries in coherent texts, allowing a separation of such effects from message level effects. Critical words occur early in the second sentence, which affords us the best chance of capturing lexical level effects due to reduced contextual constraint (Kutas et al., 1988).

2.0 Current Experiment

Our approach was to examine the effect of word association strength (taken from adult association norms; Nelson, McEvoy, & Schreiber, 1998) in both text comprehension and word meaning judgments. The word meaning judgments allow examination of the effect of association strength in a simple word-level task that requires meaning retrieval. In the text comprehension task, participants read two-sentence passages with ERPs recorded on a critical word in the second sentence, which was always the first content word of that sentence. A key comparison is between the two levels of association strength within the paraphrase condition, which linked the critical word of the second sentence to an event referred to in the first sentence by either a high-strength (*strongly associated* condition) or low-strength (*weakly associated* condition) associate of the critical word.

The paraphrase condition consisted of words that are coreferential in the given context, rather than words that are synonyms. In some cases, the second word can be said to add a meaning feature to the antecedent (e.g., “rain” adds a feature to a “storm” event), thus updating the situation model by a small increment. These conditions were compared to a *baseline* condition in which the critical word had no readily-available referent in the first sentence. Importantly, the baseline texts were not anomalous, but fully sensible and coherent, so that any differences in ERP indices of word-to-text integration could be taken to reflect an ease of integrating the critical word into the on-going situation model of the text. To allow for specific comparisons of associative strength across word text comprehension and meaning judgments, the words from the text comprehension task were also used in the word meaning task, although participants saw different word pairs.

The importance of word-to-text integration processes in comprehension implies that individual differences in these processes could explain some of the variability in previous findings, and in global off-line assessments of reading comprehension skill. Yang, Perfetti, & Schmalhofer (2005) found that the ERP patterns of less skilled adult readers were different from what Yang et al. (2007) observed for skilled readers. Whereas skilled comprehenders showed an effect of paraphrase in the typical range of the N400, this effect was not reliably observed among less skilled comprehenders. This difference led to the suggestion that word-to-text integration processes were less fluid for less skilled comprehenders. Such a difference might reflect the availability of high quality lexical representations that are hypothesized to be necessary for the retrieval of context-relevant semantics (Perfetti & Hart, 2002, Perfetti, 2007). Additionally, recent research by Andrews and Reynolds (2013) suggests that individuals with greater vocabulary knowledge are less influenced by message-level factors. Thus, a secondary goal of this study was to include assessments of readers’ comprehension ability and vocabulary knowledge to further examine individual differences in word-to-text integration processes.

In summary, the present study measured ERPs from participants with a range of reading skills during two tasks, a passive text comprehension task and a word meaning judgment task. By varying the association strength between an antecedent word in the first sentence and the critical word in a second sentence, the text comprehension task aimed to investigate the source of the paraphrase effect (Yang et al., 2007) - a reduced N400 over parietal electrodes elicited by critical words. If the effect is primarily driven by message level factors, we expected to find a reduced negativity for both the strongly associated and weakly associated text conditions relative to the baseline condition. If the effect reflects a combination of lexical level and message level integration processes, we expected to find a difference between association conditions, with critical words eliciting the least negativity in the strongly associated condition, followed by a greater negativity in the weakly associated condition, and finally, the greatest negativity in the baseline condition. In the word meaning task, we expected to replicate previous findings of graded behavioral and electrophysiological effects of association strength on the word pairs. This would confirm the efficacy of our association strength manipulation and allow for a conceptual comparison between effects in word-level and text-level on-line processing tasks among the same individuals. Finally, we examined relations between ERP measures on both tasks with offline assessments of reading skill and vocabulary knowledge.

2.1 Methods

2.1.1 Participants—Forty-one participants were recruited from the University of Pittsburgh student and staff population and paid \$10 for their participation. Some were recruited from the Pittsburgh Adult Reading Database, which includes standardized vocabulary and comprehension scores (Nelson & Denny, 1973). Other participants were recruited through advertisements placed in campus locations. The participants were right-handed, native English speakers with normal or corrected-to-normal vision, and without any history of head injury or epilepsy. The data for two participants were lost due to technical problems, and another participant's data were discarded because of an age difference (~ 40 years older than the sample mean age). Thus, data from 38 participants (Mean Years of Age (SD) = 21.8 (4.6); 23 Women) were available for further processing and analyses. After EEG pre-processing one additional participant's data for the word meaning task were removed (see below). Thus, a final sample of 38 participants contributed text comprehension data, and 37 participants contributed word meaning data. All procedures were approved by the University of Pittsburgh Institutional Review Board.

2.1.2 Materials: Text Comprehension Experiment—Word stimuli were chosen from the South Florida Association Norms database (Nelson et al., 1998). Triplets of words were chosen such that two antecedent words were available as forward associated primes of each critical word. One antecedent-critical word pair was the *strongly associated* (SA) pair, with forward association strength of at least .100 ($M = .317$). A second antecedent was chosen for the *weakly associated* (WA) pair, defined as having an antecedent-to-critical word association strength of no more than .065 ($M = .023$). These strongly and weakly associated values are comparable to those used in previous studies (e.g., Frishkoff, 2007). The mean word association strength difference of .30 between the strongly and weakly associated antecedents and critical words was significant ($p < .001$).

Backward association strength was allowed to vary freely, with the limitation that no strongly forward associated pair was of weaker backward association than its weakly forward associated complement. (This was to assure that, if backward associations are more functional in text integration, there would be no risk of their cancelling out forward effects¹.) Thus, the SA pairs had somewhat higher mean backward association strength ($M = .137$) than the WA pairs ($M = .011$), $p < .001$. Across all pairs, the correlation between forward association strength and backward association strength was $r = .396$, $p < .01$. Critical words were of medium-to-high written frequency (mean log frequency per million = 3.85; Zeno, Ivens, Millard, & Duvvuri, 1995), and average length (mean # of letters = 5.04). The strongly associated antecedents did not significantly differ from weakly associated antecedents in mean length (SA = 5.82 letters, WA = 5.44 letters; $p > .1$), or mean log frequency (SA Freq = 2.78, WA Freq = 3.07; $p > .2$).

A total of 90 groups of two-sentence passages were adapted from Yang et al. (2007) or created for this study, with each group consisting of three conditions (Table 1). The

¹Additionally, we compared latent semantic analysis cosine values (<http://lsa.colorado.edu/>) between association conditions to obtain an order-free measure of semantic similarity between word pairs. As with forward and backward association strength, the mean LSA values were greater for the SA pairs (.409) than the WA pairs (.258), a difference that was significant; $p < .001$.

associated word pairs were embedded into the passages, such that the antecedent word was in the first sentence, and the critical word (a strong or weak associate of the antecedent) was the first content word of the second sentence; in context, the critical words were conceptually related paraphrases of the antecedent words. The antecedent and critical words were separated by three to five words, as well as the sentence boundary. Finally, in the baseline (BL) condition the critical words had no antecedent in the first sentence. The BL texts were sensible, not anomalous, and they shared much the same semantic content as the paraphrase conditions. Their main difference from the SA and WA sentences was that they lacked the antecedent term of the first sentence. The three versions of the 90 experimental sentences were balanced by a Latin-square design in which each version of a passage was assigned to one of three lists. Each participant was exposed to 30 passages of each condition without repetition of any passage.

2.1.2.1. Cloze Probability Norms: The three lists created for the experiment were used, with the text passages truncated immediately prior to the critical words. Sixty Native English speakers living in the United States were recruited using Amazon's Mechanical Turk (<https://www.mturk.com/>) crowdsourcing platform. Participants were instructed to read each incomplete narrative passage for comprehension, and attempt to provide a meaningful, short (3–5 word) continuation. Each participant completed only one list, resulting in each item receiving 20 responses.

Means and standard deviations of cloze probabilities of the experimental passages are shown in Table 2. Similar patterns of responses were revealed with two different measures of cloze probability: *First-Word probability*, the probability that the first word in a response will be the critical word, and *Total probability*, the probability that one of the first 5 words in a response will contain the critical word. For both measures of predictability, the associated conditions (SA & WA) had greater average cloze probabilities than the baseline condition ($p < .01$), but did not differ in cloze probability from each other ($p > .5$).

2.1.3 Materials: Word Meaning Experiment—The word pairs collected for the text comprehension task were also used in the word meaning task. To reduce familiarity effects on the ERP responses (Molfese & Molfese, 1997), the word pairings were changed so that participants saw the critical word paired with a different antecedent word. Thus, a critical word paired with a strong associate in the text task was paired with its weak associate or with an unrelated word in the word meaning task. For the unrelated pairing, a new set of 90 unrelated words was developed. Thus, every critical word appeared as the second of a two word pair with one of three “antecedents”: strongly associated (SA), weakly associated (WA), and unrelated (Unrl). In addition, 30 unrelated filler (Filler) pairs were included so that the number of unrelated responses would equal the number of related responses. Each list contained 60 pairs of words that were related (30 SA, 30 WA), and 60 pairs of words that were unrelated (30 Unrl, 30 Filler), for 120 items total (Table 3). The unrelated “antecedent” words were matched approximately with the related “antecedent” words in length and frequency.

2.1.4 Procedure—Participants were fitted with an electroencephalogram (EEG) net and seated inside a soundproof, electrically insulated booth on an adjustable chair 60 cm from

the center of a 15-in (38.1 cm) CRT display. The text comprehension task, which was of primary interest, was first to avoid any prior exposure to the words. Participants were instructed to read the two-sentence passages for comprehension. Words appeared one at a time in black font on a white background in the center of the screen for 300 ms with an inter-stimulus interval (ISI) of 300 ms (thus a stimulus-onset asynchrony (SOA) of 600 ms). The ISI was 600ms prior to the first word of the second sentence, allowing additional time for sentence wrap-up effects (Just & Carpenter, 1980). Each two-sentence text was preceded by a fixation cross (+) in the center of the screen. A true-false comprehension question based on the meaning of the passage followed 30% of the trials on a random basis. Comprehension questions were equally divided between true and false correct responses. Their purpose was to encourage participants to read for comprehension, and immediate feedback was displayed on the screen following the responses (“Wrong” in red for incorrect responses and “Good Job” in blue for correct responses). The text task was presented in three blocks of 30 randomly ordered trials each in order to allow participants to rest, and took approximately 35 minutes to complete.

Following the text experiment and a short break, participants remained in the booth to take the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999), with their responses audio recorded for off-line scoring. The TOWRE consists of tests of word reading fluency and non-word decoding efficiency. In the word reading test participants were asked to orally read as many words as they could in 45 seconds from a sheet of paper consisting of 4 columns of words that increase in length and complexity (number of items = 104). In the non-word decoding test participants were asked to orally decode as many non-words as they could in 45 seconds from a sheet of paper consisting of three columns of non-words increasing in length and complexity (number of items = 63).

After the TOWRE was administered, participants remained in the EEG booth and took a short break prior to the word meaning task. During this task pairs of words were presented on the screen one word at a time, and, upon presentation of the second word, participants made a button-press response indicating their judgment of whether the two words were related in meaning (right index finger for ‘related’ response and right middle finger for ‘unrelated’). Each trial began with a central fixation cross (+) for 450ms, followed by a blank screen for a random duration between 75–250 ms prior to the first word, which was presented for 1000 ms and followed immediately by the second word, which was displayed for 2000 ms. After 6 practice trials with feedback, the experimental trials followed without feedback, except that a “No response” display was triggered if a response did not occur within 2000 ms. The word meaning trials were presented in 3 blocks to allow for breaks, and the stimuli were randomly presented. The task took approximately 12 minutes to complete.

This was the end of the experimental session for participants drawn from the Pittsburgh Adult Reading Database. Participants who had not taken the Nelson-Denny vocabulary and comprehension tests ended the session by taking these tests. The Nelson-Denny vocabulary test has 100 multiple-choice vocabulary items. The Nelson-Denny comprehension test features 6 passages of text followed by questions (number of questions = 36). In accord with standard procedures, 7.5 and 15 minutes were allowed for completion of the vocabulary and comprehension test, respectively.

2.1.5 ERP Recordings and Pre-Processing—ERP data were recorded using a vertex reference from a 128 electrode Geodesic sensor net (Tucker, 1993) with Ag/AgCl electrodes (Electrical Geodesics, Inc., Eugene, OR). Electrode impedances were kept below 40k Ω (Ferree, Luu, Russell, & Tucker, 2001). Six eye channels were monitored to allow for rejection of ocular artifacts. The EEG signals were digitally sampled at a rate of 500 Hz, and hardware filtered during recording between 0.1 and 200 Hz.

Data were first passed through a 30 Hz low-pass filter. Next, data for both the text comprehension trials and word meaning trials were segmented from 150 ms prior to critical words (i.e., first content word of the second sentence in the text comprehension trials, and the second word in the word meaning trials) to 600 ms after, yielding segments 750 ms long. Ocular artifact detection implemented in NetStation, based upon the regression technique of Gratton, Coles, and Donchin (1983), used the superior right eye channel to detect and regress out eyeblinks and the right outer canthi channel to detect and regress out eye-movements. Channels were considered too noisy for use and automatically removed if they had activity of ± 200 μ V on more than 20% of trials. Additionally, segments containing artifacts were removed on the basis of three separate criteria: more than 12 channels marked noisy for a given trial using the previous noisy channel thresholding step, the detection of blinks through examination of superior and inferior eye channels (excepting for the right superior eye channel removed during ocular artifact detection) for voltage fluctuations of ± 140 μ V, and the detection of horizontal eye movements (e.g., saccades) through examination of the left outer canthi electrode for voltage fluctuations of ± 55 μ V.

These procedures resulted in excessive segment loss (defined as 10 trials per condition) for one participant's word meaning task. These data were discarded. For the retained datasets, an average of 7 (SD = 1.86) electrodes was removed from the text comprehension datasets, and an average of 8.7 (SD = 2.29) was removed from the word meaning judgment datasets. Removed channels were replaced by data from neighboring channels using spherical spline interpolation (Ferree, 2006). The data was then re-referenced to the average of the channels and corrected for the Polar Average Referencing Effect (PARE; Junghöfer, Elbert, Tucker, & Braun, 1999). Average referencing is commonly used with dense-array electrode nets, and does not suffer from topographical issues seen when using average reference with sparse electrode nets (Dien, 1998; Luck, 2005). The data were then averaged for each participant for each condition. Following subtraction of the mean amplitude of the baseline period (150 ms pre-stimulus for both tasks), the data were exported to SPSS 19.0 for statistical analyses.

3.0 Results

3.1 Descriptive Skill Measures

The measures of reading and vocabulary show a wide range of scores: 12 to 35 ($M = 22.7$; $SD = 6.1$) correct on the Nelson-Denny comprehension test, and 30 to 98 ($M = 60.6$; $SD = 17.6$) on the Nelson-Denny vocabulary test. To calibrate these scores with the means of the larger population from which the participants were drawn, we compared them with 6328 participants in the Pittsburgh Adult Reading Database. The means for the present study sample were within one standard deviation of those from the larger population on both comprehension ($M = 20.9$; $SD = 5.9$) and vocabulary ($M = 49.1$; $SD = 15.6$). Participants'

scores on the TOWRE tests were just above the standard score mean of 100, both in terms of word reading ($M = 102.2$; $SD = 9.1$) and non-word decoding ($M = 105.9$; $SD = 10.4$).

Most reading measures were significantly correlated with each other, except for the zero level correlation between TOWRE Word Reading and Nelson-Denny Vocabulary scores ($r = .230$, $p = .171$). Comprehension was positively correlated with vocabulary ($r = .723$, $p < .001$), and TOWRE word reading ($r = .444$, $p = .006$) and decoding ($r = .576$, $p < .001$). Additionally, vocabulary was positively correlated with TOWRE decoding ($r = .568$, $p < .001$), and the TOWRE measures were positively correlated with each other ($r = .407$, $p = .012$). These zero-order correlations suggest that the overall picture is one of shared variance; this is also supported by the significant correlation between vocabulary and comprehension with the TOWRE measures partialled out ($r = .613$, $p < .001$).

3.2 Text Comprehension

3.2.1 Behavioral data—The comprehension questions, which were presented randomly after 1/3 of the text passages ($n = 30$), were intended to encourage participants to read the texts for comprehension. The results confirm a high level of accuracy, averaging over 90% across conditions.

3.2.2 ERP data—The major aim in the text comprehension trials was to test for message-level effects (comparing the SA and WA with BL) and lexical level effects (comparing SA with WA) within the N400 time window, when the paraphrase effect was previously found. Topographic difference maps from 300–600ms after critical word onset are shown in Figure 1.

To examine N400 differences across conditions in the text comprehension trials, mean amplitudes between 300–600 ms from the onset of the critical word were averaged in three parietal clusters (Left/P3, Central/PZ, Right/P4; figure 2), each with 7–9 electrodes. These clusters cover a large centro-parietal scalp area where the N400 effect is apparent across tasks, and was seen in previous work (Yang et al., 2007). A 3 (Condition) $\times$ 3 (Cluster) repeated-measure analysis of variance (ANOVA) was carried out on the amplitude data. This analysis revealed a reliable main effect of Condition, $F(2,74) = 3.489$, $p = .036$, $\eta_p^2 = .086$, and Cluster, $F(2,74) = 24.185$, $p < .001$, $\eta_p^2 = .395$, but no significant Condition $\times$ Cluster interaction, $F < 1$. Thus, we conducted a priori comparisons on the ERPs combined across clusters, which revealed differences between the Association Conditions and the Baseline condition. Paired comparisons showed differences between the SA and BL conditions, $t(37) = 2.115$, $p = .041$ and between the WA and BL conditions, $t(37) = 2.376$, $p = .023$, but not between the SA and WA conditions, $t(37) < 1$. These effects reflect a reduced negative deflection between 300–600ms (reduced N400) for the SA and WA conditions relative to the BL condition, particularly over central and left parietal electrodes (Figure 1; Figure 2).

To examine effects across levels of comprehension skill, the participants were divided into 3 groups on the basis of their comprehension scores in relation to those in the Pittsburgh Adult Reading Database (see above): one group at least a half standard deviation below the population mean, a second group at least a half standard deviation above the population mean, and a middle group. This produced three groups: Less Skilled ($n = 8$), mean

comprehension score of 15.12 (SD = 1.36), Skilled (n = 14), mean score of 20.21 (SD = 1.85), and More Skilled (n = 16) with a mean score of 29.23 (SD = 2.72). Since no Condition $\times$ Cluster interaction was seen in the above analysis, a second ANOVA was carried out with Condition as a within-subject factor and Comprehension Group as a between-subject factor. This analysis revealed a Condition $\times$ Comprehension Group interaction that approached significance $F(4, 70) = 2.345, p = .063, \eta_p^2 = .118$. As is shown in Figure 3, the more skilled group showed the main effect pattern of a reduced negativity for both SA and WA relative to baseline (Uncorrected t-tests: SA-BL: $t(15) = 2.104, p = .053$; WA-BL: $t(15) = 2.056, p = .058$; SA-WA: $t(15) < 1$), the skilled group showed no condition effects (Uncorrected t-tests: SA-BL: $t(13) < 1$; WA-BL: $t(15) < 1$; SA-WA: $t(15) = 1.264, p = .228$), and the less skilled group showed an N400 reduction only for the WA condition (Uncorrected t-tests: SA-BL: $t(13) < 1$; WA-BL: $t(15) = 1.968, p = .090$; SA-WA: $t(15) = -1.901, p = .099$). The less skilled group findings should be taken with caution, however, due to the small group size.

3.3 Word Meaning Task

3.3.1 Behavioral data—Our analysis focused on the comparison of main interest, Strongly Associated (SA) vs. Weakly Associated (WA) pairs. Participants judged SA pairs as related in meaning (97.2%) more often than WA pairs (79 %), $p < .001$. For decision times, we compared trials across SA and WA pairs that correctly produced Related (YES) decisions. Decisions were shorter in the SA (635 ms) condition than the WA (757 ms) condition, $p < .001$.

3.3.2 ERP data—As with the text comprehension trials, we examined the mean amplitudes from 300 to 600 ms² from the onset of the critical (second) word in the word meaning judgment trials. The same three parietal clusters (P3, PZ, P4; figure 4. A) were used for this analysis. Topographic difference maps from 300 – 600 ms after critical word onset are shown in Figure 1.

A 3 (Condition) $\times$ 3 (Electrode Cluster) repeated-measure ANOVA was carried out on the amplitude data³. Significant effects were found for Condition, $F(2, 72) = 10.360, p < .001, \eta_p^2 = .223$, Cluster, $F(2, 72) = 8.323, p = .001, \eta_p^2 = .188$, and the Condition $\times$ Cluster interaction, $F(4, 144) = 6.023, p = .001, \eta_p^2 = .143$. In a priori pairwise comparisons over Right (P4) parietal electrode clusters, both SA ($p < .001$) and WA ($p = .002$) produced a reduced N400 compared with Unrl, SA also showed a greater N400 reduction than WA at a lower level of reliability ($p = .067$) (See Figure 4). Over the left Parietal (P3) electrode cluster, SA and WA differed reliably, $p = .011$, while WA and Unrl showed a less reliable difference. $p = .086$. Overall, these results are consistent with the centro-right topography

²As an anonymous reviewer insightfully pointed out, the shorter average latencies in the SA condition relative to the WA condition may result in increased motor-related neural activity for the former condition. To help rule this potential confound out we replicated the analyses reported here using a 350–550 ms time-window and found substantively similar results.

³In the ERP analyses of the Word Meaning Task, comparisons are made between the Strongly Associated, Weakly Associated, and Unrelated Word Pairs using the critical words from the Sentence Comprehension trials. Filler items were not included in these analyses to maintain comparability with the text experiment and to have approximately equal numbers of trials in each condition. A Pairwise t-test revealed no difference between the Unrl and Filler conditions, $p > .9$.

most commonly associated with the N400 effect, in contrast to the more posterior and leftward shifted topography seen for the text comprehension task (Figure 1).

3.4 ERP Task Comparison

Although there are procedural differences between the text comprehension and word meaning tasks that are cause for caution in terms of direct statistical comparisons (see Discussion), we carried out an exploratory 3 (Condition) $\times$ 3 (Cluster) $\times$ 2 (Task) ANOVA. As expected from the separate analyses above, there was a significant Condition $\times$ Cluster $\times$ Task interaction, $F(4, 144) = 4.476$, $p = .003$, $\eta_p^2 = .111$. As can be seen in Figures 2 and 4, this difference reflects the contrast between the graded responses in the word meaning task (SA > WA > Unrl) and the different responses to the associated conditions relative to the baseline condition in the text comprehension task (SA = WA > BL).

3.5 Behavioral-ERP Correlations

Because of the differences in the topography of effects seen in the text comprehension and word meaning task, we examined correlations of off-line measures, behavioral measures, and ERP difference measures over the left parietal (P3 cluster) and right parietal (P4 cluster) electrodes separately. Research suggests that effects over different scalp locations in the N400 time window reflect different processes (Dien, Michelson, & Franklin, 2010; Franklin, Dien, Neely, Huber, & Waterson, 2007)

First, we examined whether the behavioral responses and ERP responses were related in the word meaning task. Over right parietal regions, the decision time differences between the WA and SA conditions were positively correlated with the N400 differences between these conditions ($r = .442$, $p = .006$); the WA-Unrl differences showed a smaller correlation with their N400 differences ($r = .318$, $p = .055$). Interestingly, the SA-Unrl difference in decision times was uncorrelated with the SA-Unrl N400 difference over the right parietal electrodes ($p > .5$); however this correlation was significant over left parietal electrodes ($r = .370$, $p = .024$). Thus, the N400 reduction was associated with meaning decisions times, indicating a brain-behavior link centered on lexico-semantic processing. The effects of association strength (SA vs WA) modulated this correlation specifically over the right hemisphere.

We also examined correlations between ERP effects across the two tasks, focusing on the SA and WA differences. The SA-WA differences over left parietal (P3) electrodes in the text task correlated positively with SA-WA differences over left parietal electrodes in the word meaning task ($r = .322$, $p = .052$); but they correlated negatively with SA-WA differences over right parietal electrodes (P4) in the word meaning task, ($r = -.471$, $p = .003$). SA-WA differences over the right parietal cluster was also negatively correlated between tasks ($r = -.275$, $p = .099$), suggesting a separate (set of) mechanism(s) indexed by the ERP differences in the two tasks.

3.6 Individual Differences

We compared off-line measures of reading ability with the ERP effects found in the text comprehension task. Over the left (P3) Parietal cluster, vocabulary correlated with the difference between the mean amplitudes of SA and WA conditions ($r = .320$, $p = .05$), with

higher vocabulary associated with relatively larger amplitudes for SA than WA trials. Comprehension was less reliably correlated with this difference ($r = .192$, $p = .247$). Additionally, TOWRE word reading was positively correlated with the difference between the mean amplitudes of the SA and BL conditions ($r = .389$, $p = .017$).

In the word meaning task, SA-WA N400 differences showed negative correlation with vocabulary over right parietal ($r = -.304$, $p = .067$), and to a lesser extent over left parietal ($r = -.034$, $p = .840$) regions. This suggests that the ERP meaning effect for participants with larger vocabularies was less affected by association strength over the area that is typically associated with the N400.

4.0 Discussion

We examined the influence of message level (coreference) and lexical level (word association strength) factors on the on-line processing of single words during passive reading of short texts and during word meaning judgments. We also examined the relation between ERPs and behavioral measures across these two tasks. Finally, we tested the correlation of global, offline measures of reading ability with ERP responses to the critical contrasts in these tasks.

A key result of the text comprehension task is that an ERP indicator of word-to-text integration, the N400, was observed across a sentence boundary when a single critical word referred to an event that was established by a word in the previous sentence. This paraphrase effect, replicating Yang et al. (2007), did not depend on the associative strength between the word in the first sentence and the critical word in the second sentence. Thus, within the associative strength parameters of the study, our results indicate that message level effects—the use of word meaning to construct connections that depend on the text meaning—are primarily responsible for the integration of referential meaning across sentence boundaries. This functioning of message level factors in text comprehension was reflected in a more central to left posterior distribution of N400 effects than those seen in studies using simpler materials (Kutas & Hillyard, 1980), or those with strong contrasts in terms of predictability (for review see Federmeier, 2007) or congruence (Coulson et al., 2005), where the demands on message level skills may be minimized. The predictability of our materials was relatively low, as can be seen in Table 2, and, importantly, critical words in the strongly and weakly associated texts were equally predictable. Although the critical words in the associated conditions were slightly more predictable than critical words in the baseline condition, all conditions were created to be congruent and without strong predictability for an individual word, leading to greater pressure on message level processing. These considerations support a message level interpretation of the N400 effects from our task comprehension task.

The message level processing is most in evidence for those readers who had high global comprehension ability. The most skilled group showed a pattern of N400 reduction for both strongly and weakly associated words, whereas for the small group of least skilled comprehenders this effect was only for weakly associated words (Figure 3). This global comprehension skill depends on a mix of word knowledge and word processing with higher-level language skills, e.g., the ability to monitor the coherence of texts (Baker, 1984; Garner,

1980), to draw inferences (Oakhill & Garnham, 1988), and domain general processes, e.g. working memory (Just & Carpenter, 1992; but see Van Dyke, Johns, & Kukona, 2014). The present study included assessments of word reading skill, vocabulary knowledge, and general comprehension skill. The strong correlation between comprehension and vocabulary skills in our sample attests to the shared variance between these two reading-related abilities, even in relatively skilled adult readers.

We have used the phrase “word-to-text integration” to describe the process of referential binding between the currently read word and a previously introduced word or phrase. The corresponding interpretation of the N400 in text experiments is that it reflects, among other possibilities, integrative processes that occur with meaning congruence. However, other interpretations of the N400 in text processing have been proposed, including that it reflects lexical access (Lau et al., 2009; Lau, Phillips, & Poeppel, 2008) and is the result of predictive processes (Federmeier, 2007). Although prediction can be reflected in the N400 and is relevant for text comprehension, integration processes are needed in comprehension at the level of the situation model (Kintsch, 2008) and the N400 can reflect these processes along with those that are necessary for integration (lexical access) or can support it (prediction). Indeed, it has been suggested that the N400, and the N400 effects seen in psycholinguistic research, are the result of two separate processes akin to semantic activation and integration/unification (Baggio & Hagoort, 2011). The contribution of integrative processes to the generation of the N400 component the N400 component is supported by findings that sentences containing complement coercions (e.g., “The journalist began the article”; Baggio, Choma, van Lambalgen, & Hagoort, 2010). Processing these structures, which requires additional integrative or combinatorial processes, leads to larger N400 responses to the coercing noun compared with non-coercing and semantically congruent control nouns (Baggio et al., 2010; Kuperberg, Choi, Cohn, Paczynski, & Jackendoff, 2010). Within this framework, the reduced N400s in the associatively paraphrased texts conditions compared to the baseline condition can be taken to index the relative ease with which the critical word is bound to its preceding coreferential context. This may be particularly true of materials like ours, which are of low predictability, and thus may be more naturalistic than extremely constraining materials.

In contrast to the text reading results, the same key words in a word meaning task showed clear effects of associative strength. Behaviorally, when subjects decided that two words were related in meaning, strongly associated words produced more rapid meaning decisions than did words that were weakly associated. In the ERP data, a graded response to associative strength was found over right parietal electrodes in the N400 time window. Words that were strongly associated showed a reduced negativity compared with both unrelated words and weakly associated words, and weakly associated words showed a reduced negativity relative to unrelated words. Both the behavioral and ERP findings from the word meaning task are consistent with research showing a facilitative effect of word association on reaction time (Coney, 2002) and ERP indices of lexico-semantic processing (Holcomb, 1988; Kutas & Hillyard 1984). Thus, the same difference in associative strength that failed to modulate the ERP indicator of the paraphrase effect in text comprehension did modulate it in the word meaning task, a finding supported by a statistical comparison that included task as a factor. In contrast to the text comprehension findings, the largest N400

effects in the word meaning task were seen over right parietal locations. All conditions differed in the direction expected, consistent in time window and topography with the typical N400 (See Kutas & Federmeier, 2011).

The difference in patterns across the two tasks partially reflects the processing demands they place on lexico-semantic processes. Whereas word meaning judgments require the retrieval and comparison of lexical meanings, text comprehension has the additional requirement of relating the currently read word to the understanding of the text that already has been achieved, a requirement that increases the need for semantic integration/unification processes (Baggio & Hagoort, 2008; Hagoort, 2005; 2013). At a minimum, this requires a word's meaning to be integrated with a linguistic (propositional) representation of the text; at a deeper level, the linkage is to a referential representation of the text meaning (the situation model). These task-related differences may be reflected in ERP patterns evoked by a single word.

In our case, the two tasks showed differences in the effect of association strength and also in the scalp locations of these association strength effects. Some research has suggested that ERPs recorded at different scalp sites in the same time window may index different processes (Dien, Michelson, & Franklin, 2010), for example, semantic priming over left parietal sites (Dien & O'Hara, 2008), and semantic matching over right parietal electrodes (Franklin et al., 2007). The differential topography for SA-WA effects in the text comprehension task, relative to the word task, may reflect the differential activation of neural systems involved in higher-level language and cognition (Fedorenko, Behr, & Kanwisher, 2011; Mason & Just, 2004; Perfetti & Frishkoff, 2008). One might question whether differences between the tasks make comparisons problematic. There were slightly longer prime-target stimulus-onset asynchronies (SOAs) in the word meaning judgments (i.e., 1000ms) relative to the SOAs between words in the texts (i.e., 600 ms). Also, explicit responding was required in the word judgment task, while the point of ERPs in reading is to detect processes without explicit responding. Certainly such differences should be kept in mind. However, the two tasks were designed to share the critical words in order to allow a benchmark for effects in text comprehension, i.e. that word pairs used in the text condition actually produce N400s that are affected by association strength when word meaning retrieval was explicitly involved. This indeed was the result.

The largest correlation between off-line reading measures and the ERP measures during text reading was that of vocabulary and the association strength difference (SA-WA) observed over the left parietal scalp. Comprehension skill, which showed a smaller correlation with this effect, is relevant for processing all the experimental texts and apparently was not differentially important for those texts that had coreferential ties. Word meaning knowledge would also be important for all the experimental texts, but showed a special relevance for the texts when strongly associated words were involved in the coreferential integration process. In the word meaning task, in contrast, the correlation between vocabulary ability and association strength differences (SA-WA) was negative over right parietal scalp. Thus, in a simple word meaning task, higher vocabulary was associated with a reduced effect of association strength over typical N400 sites; however, in text comprehension, higher vocabulary was associated with a greater effect of association over left hemisphere sites that

have been associated with greater semantic priming (Dien & O'Hara, 2008). These findings suggest complementary effects of association strength and word meaning knowledge. Such knowledge allows priming to occur, possibly affecting the early stages of text integration; but when the focus of the task is the explicit comparison of single word meanings, knowing more word meanings eliminates the advantage of strong association. For lower vocabulary participants, their weaker knowledge of some word meanings produces an advantage for words that are strongly associated.

In considering differences in overall comprehension skill, it is important to note the variance it shares with word meaning (vocabulary) knowledge. Our reading task measurement targeted word reading in text, exposing the word meaning processes in comprehension. The contrast between the most and least skilled comprehenders reflects the abilities of the most skilled group to retrieve the meaning of a word and interpret it in a way that is sensitive to the message context; thus there was no effect of association strength. By contrast, the least skilled comprehenders showed the N400 integration effect only for weak associations. This contrast implicates individual differences in processes that establish text meaning, including the processes that retrieve and integrate word meanings. However, the small number of participants in the least skilled group (8) is a reminder (as is the lack of any clear pattern of condition effects for the middle group of comprehenders) that more is needed to test whether there is a general relationship between comprehension skill and the dependence of the N400 on word association strength.

There are two considerations in interpreting the paraphrase effect. First, our results concerned cross-sentence effects only. It is possible that message level effects (relative to lexical effects) emerge most strongly when sentence and clause boundaries mark the end of one or more full verb phrases. Longer eye fixations on sentence-final words (Just & Carpenter, 1980; Rayner et al., 1989) and poorer verbatim memory across sentence boundaries in reading (Goldman, Hogaboam, Bell, & Perfetti, 1980) imply that word-specific information becomes less available across sentence boundaries. A possible refinement of these boundary effects is that comprehension processes are managed by clausal and sentence boundaries in a way that strengthens the message level representation at the expense of the lexical level.

A second qualification concerns the function and direction of lexical associations. Our study focused on the contrast between strong and weak *forward* association in short text passages. However, it may be *backward* associations that are most relevant for text processing. Comprehension relies on memory processes that make text representations — words, propositions, and various dimensions of structure (Graesser, Singer, & Trabasso, 1994; Zwaan, Langston, & Graesser, 1995; Zwaan & Radvansky, 1998) — available as a word is read and then integrated into the reader's understanding of the text. These memory processes can be described by a resonance mechanism that automatically links what is read to what is accessible in memory (O'Brien, Rizzella, Albrecht, & Halleran, 1998). Integrative processes may be facilitated when an encountered word activates its associations, which then resonate with the memory of a word recently encountered in the text. If so, backward association strength, which would capture some of this process, may turn out to be more important than forward association strength. In a recent study completed in our lab, ERP effects of

backward association between words across sentence boundaries were seen earlier than effects of forward association, and effects of association direction differed depending on scalp location during the N400 time window (Stafura, Perfetti, & Rickles, In Review).

The present study adds to the previous research showing discourse effects in comprehension in the following specific ways: First, the results are the first to show N400 effects at the first content word across a sentence boundary. Words early in sentences elicit larger N400 responses than those later in sentences (Kutas, Van Petten, & Besson, 1988), and larger N400 responses to low frequency words compared to high frequency words are only apparent when the words are in sentence initial positions (Van Petten & Kutas, 1990). These findings point to the limited contextual constraint available early on in sentences.

Additionally, the effects we observed reflected not the difference between congruence and incongruence but the differences due to the availability of co-referencing. The lack of associative strength effects suggests that the effects are driven by the discourse level, consistent with studies using different approaches; in particular, with van Berkum et al (1999), whose study made sentence meaning incompatible with discourse meaning measured ERPs on words at or near the end of sentences. Our study shows effects that are localized on a word at a point at which sentence meaning has not been established, and is consistent with the hypothesis that the N400 in discourse processing can reflect lexically based coreferential processes as well as expectancy violations (Baggio & Hagoort, 2011).

In broader theoretical terms, our study joins others (Anderson & Holcomb, 2005; Otten & van Berkum, 2008; van Berkum et al., 1999; Yang et al., 2007) in exposing the intimate connection between word processes and the reader's momentary representation of text meaning. These processes are not exclusively about the expectation of specific words, but include relating a word's meaning, as it is read, to text meaning. The stronger interpretation, which is yet to be demonstrated, is that these immediate integration processes operate in part on the level of a referential mental model, rather than exclusively on linguistic representations. The theoretical question of interest is then whether the semantic coding of the text is propositional, in which either prediction or memory is about words and phrases, or nonpropositional, in which the meaning representations are highly referential.

At a general level the findings demonstrate some of the different ways word meanings are used. The importance of stored word meanings and associative connections is seen clearly in the simple task of comparing the meanings of two words. In text reading, the stored meaning of a word and its associations continue to play a role, but greater role for context-relevant (message level) word processing emerges. This allows word-to-text integration to occur, not solely on the basis of a word's stored lexical meaning and associative connections, but through flexible use of a word's meaning to integrate with referents already established by the text and the reader's situation model.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported in part by NICHD grant R01HD058566-02 to Charles A. Perfetti. The authors are grateful to Suzanne Adlof, Laura Halderman, and Ben Rickles for insightful discussion throughout the preparation and completion of this study, and to Jen Dandy, Kim Muth, and Jon-Michel Seman for help in carrying out experimental sessions.

References

- Anderson JE, Holcomb PJ. An electrophysiological investigation of the effects of coreference on word repetition and synonymy. *Brain and language*. 2005; 94(2):200–216. [PubMed: 15896394]
- Andrews, S.; Reynolds, G. Why it is easier to wreak havoc than unleash havoc. In: Britt, A.; Goldman, SR.; Rouet, JF., editors. *From words to reading for understanding*. New York, NY: Routledge; 2013. p. 72–91.
- Baggio G, Choma T, van Lambalgen M, Hagoort P. Coercion and compositionality. *Journal of Cognitive Neuroscience*. 2010; 22:2131–2140. [PubMed: 19583469]
- Baggio G, Hagoort P. The balance between memory and unification in semantics: a dynamic account of the N400. *Language and Cognitive Processes*. 2011; 26(9):1338–1367.
- Baggio G, van Lambalgen M, Hagoort P. Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language*. 2008; 59(1):36–53.
- Baker L. Spontaneous versus instructed use of multiple standards for evaluating comprehension: Effects of age, reading proficiency, and type of standard. *Journal of Experimental Child Psychology*. 1984; 38:289–311.
- Balota DA, Lorch R. Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1986; 12:336–345.
- Bentin S, McCarthy G, Wood CC. Event-related potentials associated with semantic priming. *Electroencephalography and Clinical Neurophysiology*. 1985; 60:343–355. [PubMed: 2579801]
- Camblin CC, Gordon PC, Swaab TY. The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*. 2007; 56(1):103–128. [PubMed: 17218992]
- Carroll P, Slowiaczek ML. Constraints on semantic priming in reading: A fixation time analysis. *Memory & Cognition*. 1986; 14(6):509–522. [PubMed: 3796288]
- Coney J. The effect of associative strength on priming in the cerebral hemispheres. *Brain and Cognition*. 2002; 50:234–241. [PubMed: 12464192]
- Coulson S, Federmeier KD, Van Petten C, Kutas M. Right hemisphere sensitivity to word- and sentence-level context: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31(1):129–147.
- Dien J. Issues in the application of the average reference: Review, critiques, and recommendations. *Behavioral Research Methods*. 1998; 30:34–43.
- Dien J, Michelson CA, Franklin MS. Separating the visual sentence N400 effect from the P400 sequential expectancy effect: Cognitive and neuroanatomical implications. *Brain Research*. 2010; 1355:126–140. [PubMed: 20691670]
- Dien J, O'Hara AJ. Evidence for automatic sentence priming in the fusiform semantic area: Convergent ERP and fMRI findings. *Brain Research*. 2008; 1243:134–145. [PubMed: 18840418]
- Federmeier KD. Thinking ahead: The roots and roles of prediction in language comprehension. *Psychophysiology*. 2007; 44:491–505. [PubMed: 17521377]
- Federmeier KD, Kutas M. A rose by any other name: Long-term memory structure and sentence processing. *Journal of memory and Language*. 1999; 41(4):469–495.
- Fedorenko E, Behr MK, Kanwisher N. Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*. 2011; 108(39):16428–16433.
- Ferree TC. Spherical splines and average referencing in scalp electroencephalography. *Brain Topography*. 2006; 19(1–2):43–52. [PubMed: 17019635]

- Ferree TC, Luu P, Russell GS, Tucker DM. Scalp electrode impedance, infection risk, and EEG data quality. *Journal of Clinical Neurophysiology*. 2001; 112:536–544.
- Franklin MS, Dien J, Neely JH, Huber E, Waterson LD. Semantic priming modulates the N400, N300, and N400RP. *Clinical Neurophysiology*. 2007; 118:1053–1068. [PubMed: 17336145]
- Frishkoff GA. Hemispheric differences in strong versus weak semantic priming: Evidence from event-related brain potentials. *Brain and Language*. 2007; 100:23–43. [PubMed: 16908058]
- Garner R. Monitoring of understanding: An investigation of good and poor readers' awareness of induced miscomprehension of text. *Journal of Reading Behavior*. 1980; 12:55–63.
- Goldman SR, Hogaboam TW, Bell LC, Perfetti CA. Short-term retention of discourse during reading. *Journal of Educational Psychology*. 1980; 72:647–655.
- Gouldthorp B, Coney J. Integration and coarse coding: Right hemisphere processing of message-level contextual information. *Laterality*. 2011; 16(1):1–23. [PubMed: 19657950]
- Graesser AC, Singer M, Trabasso T. Constructing inferences during narrative text comprehension. *Psychological Review*. 1994; 101(3):371–395. [PubMed: 7938337]
- Gratton G, Coles MGH, Donchin E. A new method for off-line removal of ocular artifacts. *Electroencephalography and Clinical Neurophysiology*. 1983; 55:468–484. [PubMed: 6187540]
- Hagoort P. MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*. 2013; 4:416. [PubMed: 23874313]
- Hagoort P. On Broca, brain, and binding: a new framework. *Trends in cognitive sciences*. 2005; 9(9): 416–423. [PubMed: 16054419]
- Holcomb P. Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*. 1988; 35:66–85. [PubMed: 3179703]
- Johnson-Laird PN. Mental models in cognitive science. *Cognitive science*. 1980; 4(1):71–115.
- Junghöfer M, Elbert T, Tucker DM, Braun C. The polar average reference effect: A bias in estimating the head surface integral in EEG recording. *Clinical Neurophysiology*. 1999; 110:1149–1155. [PubMed: 10402104]
- Just MA, Carpenter PA. A theory of reading: From eye fixations to comprehension. *Psychological Review*. 1980; 87(4):329–354. [PubMed: 7413885]
- Just MA, Carpenter PA. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*. 1992; 99:122–149. [PubMed: 1546114]
- Kintsch W. The role of knowledge in discourse comprehension: a construction integration model. *Psychological Review*. 1998; 95(2):163. [PubMed: 3375398]
- Kuperberg G, Choi A, Cohn N, Paczynski M, Jackendoff R. Electrophysiological correlates of complement coercion. *Journal of Cognitive Neuroscience*. 2010; 22:2685–2701. [PubMed: 19702471]
- Kutas M, Federmeier KD. Thirty years and counting: Finding meaning in the N400 component of the event related potential (ERP). *Annual Review of Psychology*. 2011; 62:621–647.
- Kutas M, Hillyard SA. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*. 1980; 207:203–205. [PubMed: 7350657]
- Kutas M, Hillyard SA. Brain potentials reflect word expectancy and semantic association during reading. *Nature*. 1984; 307:161–163. [PubMed: 6690995]
- Kutas M, Van Petten C, Besson M. Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology*. 1988; 69:218–233. [PubMed: 2450003]
- Landauer TK, Dumais ST. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*. 1997; 104(2):211.
- Landauer TK, Foltz PW, Laham D. An introduction to latent semantic analysis. *Discourse processes*. 1998; 25(2–3):259–284.
- Lau E, Almeida D, Hines PC, Poeppel D. A lexical basis for N400 context effects: Evidence from MEG. *Brain & Language*. 2009; 111:161–172. [PubMed: 19815267]
- Lau E, Phillips C, Poeppel D. A cortical network for semantics: (de)Constructing the N400. *Nature Reviews Neuroscience*. 2008; 9:920–933.

- Ledoux K, Camblin CC, Swaab TY, Gordon PC. Reading words in discourse: The modulation of lexical priming effects by message-level context. *Behavioral and Cognitive Neuroscience Reviews*. 2006; 5(3):107–127. [PubMed: 16891554]
- Luck, SJ. An introduction to the event-related potential technique. Cambridge: MIT Press; 2005.
- Mason RA, Just MA. How the brain processes causal inferences in text. *Psychological Science*. 2004; 15(1):1–7. [PubMed: 14717824]
- Meyer DE, Schvaneveldt RW. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*. 1971; 90:227–234. [PubMed: 5134329]
- Molfese, D.; Molfese, D. The use of brain recordings to assess learning. In: Mead, J., editor. *Proceedings of the International Conference on Education in Engineering*. Carbondale: Southern Illinois University; 1997.
- Morris RK. Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1994; 20:92–103.
- Neely, JH. Semantic priming effects in visual word recognition: A selective review of current findings and theories. In: Besner, D.; Humphreys, G., editors. *Basic processes in reading: Visual word recognition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.; 1991. p. 264–336.
- Neely, JH.; Keefe, DE. Semantic context effects on visual word processing: a hybrid prospective/retrospective processing theory. In: Bower, GH., editor. *The psychology of learning and motivation: advances in research and theory*. New York: Academic Press; 1989. p. 207–248.
- Nelson, MJ.; Denny, EC. *The Nelson-Denny Reading Test*. Boston: Houghton Mifflin; 1973.
- Nelson DL, McEvoy CL, Schreiber TA. The University of South Florida word association, rhyme, and word fragment norms. 1998 <http://www.usf.edu/FreeAssociation/>.
- Oakhill, J.; Garnham, A. *Becoming a skilled reader*. New York: Basil Blackwell; 1988.
- O'Brien EJ, Rizzella ML, Albrecht JE, Halleran JG. Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1998; 24(5):1200–1210.
- Otten M, Van Berkum JJA. Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*. 2008; 45:464–496.
- Perfetti CA. Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*. 2007; 11(4):357–383. (2007).
- Perfetti, CA.; Frishkoff, GA. The neural bases of text and discourse processing. In: Stemmer, B.; Whitaker, HA., editors. *Handbook of the neuroscience of language*. Cambridge, MA: Elsevier; 2008. p. 165–174.
- Perfetti, CA.; Hart, L. The lexical quality hypothesis. In: Vehoeven, L.; Elbro, C.; Reitsma, P., editors. *Precursors of functional literacy*. Washington, DC: American Psychological Association; 2002. p. 189–213.
- Rayner K, Sereno SC, Morris RK, Schmauder AR, Clifton C. Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*. 1989; 4:21–49.
- Stafura JZ, Perfetti CA, Rickles B. Which way to integration? Event-related potential evidence for mechanisms of word-by-word comprehension. (In Review).
- Torgesen, JK.; Wagner, RK.; Rashotte, CA. *Test of word reading efficiency*. Austin, TX: Pro-Ed.; 1999.
- Traxler MJ, Foss DJ, Seely RE, Kaup B, Morris RK. Priming in sentence processing: Lexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research*. 2000; 29(6):581–595. [PubMed: 11196064]
- Tucker DM. Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*. 1993; 87:154–163. [PubMed: 7691542]
- Tyler LK, Marslen-Wilson WD. The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*. 1977; 16:683–692.
- Van Berkum JJ, Brown CM, Zwitserlood P, Kooijman V, Hagoort P. Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31(3):443–467.

- Author Manuscript
- Author Manuscript
- Author Manuscript
- Author Manuscript
- Van Berkum JJ, Hagoort P, Brown CM. Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*. 1999; 11(6):657–671. [PubMed: 10601747]
- Van Dyke JA, Johns CL, Kukona A. Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*. 2014; 131(3):373–403. [PubMed: 24657820]
- Van Petten C. A comparison of lexical and sentence-level effects in event-related potentials. *Language and Cognitive Processes*. 1993; 8(4):485–531.
- Van Petten C, Coulson S, Weckerly J, Federmeier KD, Folstein J, Kutas M. Lexical association and higher-level semantic context: An ERP study. *Journal of Cognitive Neuroscience Supplement*. 1999; 46
- Van Petten C, Kutas M. Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*. 1990; 18(4):380–393. [PubMed: 2381317]
- Wicha NY, Moreno EM, Kutas M. Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*. 2004; 16(7):1272–1288. [PubMed: 15453979]
- Yang CL, Perfetti CA, Schmalhofer F. Less skilled comprehenders' ERPs show sluggish word-to-text integration processes. *Written Language and Literacy*. 2005; 8:157–181.
- Yang CL, Perfetti CA, Schmalhofer F. Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2007; 33(1):55–89.
- Zeno, SM.; Ivens, SH.; Millard, RT.; Duvvuri, R. The educator's word frequency guide. New York: Touchstone Applied Science Associates (TASA), Inc.; 1995.
- Zwaan RA, Langston MC, Graesser AC. The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*. 1995; 6(5):292–297.
- Zwaan RA, Radvansky GA. Situation models in language comprehension and memory. *Psychological Review*. 1998; 123(2):162–185.

Highlights

- We examine effects of strong and weak lexical association in text comprehension
- We examine association effects in word meaning judgments for same participants
- Association strength elicits graded N400 reductions in word meaning task
- No difference between N400 in text comprehension association conditions
- Most skilled comprehenders drive text comprehension effects

Figure 1.
A. Grand average topographic voltage maps of condition differences in text comprehension (left) and meaning judgment (right) tasks from 300 to 600 ms after word onset. Voltage scale is $-2\mu\text{V}$ to $2\mu\text{V}$, with warm colors indicating positive voltages, and cold colors indicating negative voltages. SA = Strongly Associated. WA = Weakly Associated. Base = Baseline. Unrl = Unrelated.

Figure 2. Grand average waveforms and difference graph for the Text Comprehension trials. Strongly Associated is in blue, Weakly Associated is in red, and Baseline is in black. The solid vertical line on the left side of the waveform indicates stimuli onset. The dotted box on the right side of the waveforms indicates the 300–600ms region used in the N400 analysis. A. Electrode layout (anterior is at top), highlighting the three clusters of electrodes. B. Left Parietal (P3) cluster waveform. C. Centro-Parietal (Pz) cluster waveform. D. Right Parietal (P4) cluster waveform. E. Condition differences (in μV) averaged across Parietal clusters from 300–600ms after critical word onset. SA = Strongly Associated. WA = Weakly Associated. Base = Baseline.

Figure 3. Condition difference in text comprehension by off-line comprehension ability. Y-axis is in amplitude average (in μV) on the Text Comprehension trials, for all participants combined (top row), and for the different comprehension groups in the study (bottom row). SA = Strongly Associated. WA = Weakly Associated. Base = Baseline.

Figure 4. Grand average waveforms and differences graphs for the Word Meaning trials. Strongly Associated is in blue, Weakly Associated is in red, and Unrelated is in black. The solid vertical line on the left side of the waveform indicates stimuli onset. The dotted box on the right side of the waveforms indicates the 300–600ms region used in the N400 analysis. A. Electrode layout (anterior is at top), highlighting the three clusters of electrodes. B. Left Parietal (P3) cluster waveform. C. Centro-Parietal (Pz) cluster waveform. D. Right Parietal (P4) cluster waveform. E. Condition differences (in μV) averaged across the P3 cluster from 300–600ms after critical word onset. F. Condition differences (in μV) averaged across the Pz cluster from 300–600ms after critical word onset. G. Condition differences (in μV) averaged across the P4 cluster from 300–600ms after critical word onset. SA = Strongly Associated. WA = Weakly Associated. Unrl = Unrelated.

Table 1

Sample Passages for Each Experimental Condition.

Text Condition	Sample Passage
Strongly Associated Paraphrase	While Cathy was riding her bike in the park, dark clouds began to gather, and it started to <u>storm</u> . The rain ruined her beautiful sweater.
Weakly Associated Paraphrase	While Cathy was riding her bike in the park, dark clouds began to gather, and it started to <u>shower</u> . The rain ruined her beautiful sweater.
Baseline	When Cathy saw there were no dark clouds in the sky, she took her bike for a ride in the park. The rain that was predicted never occurred.

Note. The critical word (rain) is underlined and in bold at the beginning of the second sentence. The antecedent words in the paraphrase conditions are underlined.

Table 2

Cloze Probabilities for Text Comprehension Materials

Condition	First-Word Probability Mean (SD)	Total Probability Mean (SD)
Strongly Associated	.053 (.10)	.069 (.11)
Weakly Associated	.062 (.11)	.079 (.13)
Baseline	.007 (.03)	.014 (.04)

Note. First-Word probability is the probability that the first word in a cloze response will be the critical word. Total probability is the probability that one of the first 5 words in a cloze response will contain the critical word.

Table 3

Examples of Word Pair Stimuli in Experiment

Condition	Example Word Pairs	Forward Association Strength ^a	LSA Cosine ^b
Strongly Associated	Storm → Rain	.368	.56
Weakly Associated	Shower → Rain	.011	.32
Strongly Associated	Blanket → Cover	.197	.31
Weakly Associated	Sheet → Cover	.021	.20

Note.

^a According to the South Florida Association Norm Database (Nelson et al., 1998).^b Gathered from <http://lsa.colorado.edu/>.