Published in final edited form as:
Neuroimage. 2012 August 1; 62(1): 356–366. doi:10.1016/j.neuroimage.2012.04.054.

So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning

Edward W. Wlotko[a] and **Kara D. Federmeier**[b]

[a]Beckman Institute, University of Illinois, 405 N. Mathews Avenue, Urbana, IL 61801, phone: (217) 333-8864, fax: (217) 333-2922, ewwlotko@alumni.pitt.edu

[b]Beckman Institute, Department of Psychology, and Program in Neurosciences, University of Illinois, 603 E. Daniel Street MC 716, Champaign, IL 61820, kfederme@illinois.edu

Abstract

Factors that modulate the influence of contextual information on semantic processing in language comprehension have been thoroughly investigated with the N400 component of the event-related potential (ERP), a direct measure of initial contact with semantic memory. Although context has a strong and immediate impact on processing, multiple mechanisms contribute to the construction of message-level representations during normal comprehension. Some of these may be engaged after or concurrent with the formation of an initial meaning representation, and can then serve to revise or reshape meaning. In this study, ERPs were recorded while participants read plausible sentences that continuously varied in the amount of contextual constraint for the sentence-final word, defined via extensive norming data including the range of possible alternative completions for the contexts. Consistent with numerous past studies, the amplitude of the N400 was graded with expectancy, as amplitudes decreased with increasing constraint. Additionally, a left-lateralized, broad, slow negativity onsetting around 400–500 milliseconds was largest for sentences with moderately strong constraint. Within this range of constraint, the negativity was larger for sentences with fewer alternative completions compared to those with many different ones. The timing and scalp distribution of the effect resembles brain responses linked to engagement of working memory resources, ambiguity resolution, and comprehension of jokes. Similar to cases of "frame-shifting" in non-literal language, this effect may reflect processing associated with reinterpretation or reconsideration of contextual material when multiple interpretations of a sentence were likely.

Keywords

event-related potentials; language; sentential context; N400; frontal/anterior negativity; frame-shifting

Correspondence to: Edward W. Wlotko.

1 Introduction

Constructing meaning from written text or spoken utterances, although seemingly effortless, involves complex cognitive and neural processing that spans levels ranging from perception to social cognition. Consequently, the long history of inquiry into how readers and listeners understand the meaning of what they see or hear includes investigations of how single words are recognized, how words are combined into sentences, how message-level information is integrated across larger pieces of text or discourse, and how extra-linguistic information affects comprehension at all of these stages (see, e.g., Schwanenflugel & LaCount, 1988; Simpson, 1991; Van Berkum, 2008).

Within this literature, a once canonical and still prevalent view posits a hierarchical processing stream, in which individual words are initially recognized and their meaning retrieved in the same, stereotyped way regardless of the context in which the word appears (e.g. Forster, 1981). This view hypothesizes that it is only after the completion of a discrete stage of word recognition, during which context-independent meaning is accessed, that word meanings are combined to create higher level units (e.g., sentences) and that context information from the sentence and discourse level becomes available to shape word processing (e.g. Swinney, 1979; Clifton, Staub, & Rayner, 2007).

Countering this view is evidence suggesting that contextual information is available to affect word processing at the earliest points that any semantic effects can be observed (e.g. Van Petten, 1995; Van Berkum et al., 2003; Kamide, Altmann, & Haywood, 2003). One particularly important source of such evidence has come from human electrophysiology studies, which offer the ability to scrutinize comprehension as it occurs, with high temporal resolution. Because event-related brain potentials (ERPs) provide multidimensional measures, including well-studied components that have been linked to particular cognitive and neural processes, they offer the ability to examine not only quantitative but also qualitative differences in comprehension (see Kutas et al., 2007). Furthermore, the ERP technique provides its dependent measures without the need to impose any task over and above reading for comprehension, mitigating the possibility that task-dependent effects will contaminate more naturalistic comprehension goals and processes. ERP research can thus provide a fairly detailed picture of the processing engaged over time during reading or listening.

Effects of higher-level context information, including sentence-level and discourse-level information, world knowledge, and speaker characteristics (Van Petten, 1993; Kutas, 1993; Federmeier et al., 2003; Van Berkum et al., 2008) are observed in the ERP signal as early as the first word-level semantic effects (e.g., lexical associative priming). Moreover, these higher-level context effects often override the word-level ones (e.g. Coulson et al., 2005). Furthermore, factors such as message-level congruity have been shown to modulate ERP responses prior to word recognition (Van Petten et al., 1999) and to apply to stimuli that are unfamiliar and irregular (and hence cannot be recognized; Laszlo & Federmeier, 2009). Thus, semantic processing begins before all potentially relevant information is available, but seems to be modulated by whatever contextual information is currently available. Evidence like this has shifted models of comprehension away from viewing “word recognition” as marking the end of a discrete processing stage that is isolated from context (see also, e.g., McClelland, St. John, and Taraban, 1989) and toward conceptualizing recognition and meaning access as arising through a continuous accrual and synthesis of information both from the stimulus and from the broader context in which it is embedded. On such accounts, the timing with which particular types of information become available will importantly determine whether or not -- and if so, how -- they are used during on-line processing.

For example, Federmeier & Laszlo (2009) proposed that there may be a temporal constraint governing how context information can be used to shape initial semantic activations for words. Their proposal focuses on the N400, a well-studied component of the ERP that indexes access to semantic memory (see Kutas & Federmeier, 2011). All potentially meaningful stimuli across modalities, including written and spoken words, will elicit N400 activity, which follows sensory components that reflect modality-specific perceptual processes and feature extraction at varying levels of complexity (see, e.g., Holcomb & Grainger, 2007). A wide variety of manipulations affecting semantic processing of stimuli will modulate the amplitude of the N400. For example, when words are repeated, placed in the context of an associated word, or encountered in a congruent sentence context, N400 amplitudes will be reduced compared to the same stimuli out of context or in an incongruous or unsupportive context. As such, reduction in N400 amplitude can be taken as a measure of reduced demands on semantic processing. These same types of manipulations, however, do not affect the peak latency of the N400. In fact, the latency of the N400 is remarkably stable across experimental manipulations, and is generally affected only by subject-level variables such as age, reading ability, or disease state; or sometimes by task-level manipulations such as presentation rate (e.g. Kutas, 1987).

This stability led Federmeier & Laszlo (2009) to suggest that the brain activity observed in the N400 may instantiate a temporally-constrained functional “binding”, wherein semantic information accrued from context and from bottom-up processing of the eliciting stimulus are linked, through temporal synchrony, to create a stable representation. That is, in the course of normal comprehension, sentences, conversational exchanges, and discourses unfold through a series of stimuli (words) to be analyzed and processed for message-level meaning. In addition to the linguistic stream, which contains its own types of contextual constraints, interlocutors can also rely on co-present objects or scenes, gestures, and other nonverbal signals. The problem of integrating these separate pieces of relevant meaning-related information -- while still maintaining the ability to distinguish between separate stimuli -- is a binding problem that may be similar to that of fusing separate streams of features together in the visual processing of objects. Federmeier & Laszlo propose that this binding may be accomplished through timing: in this view, N400 amplitude reflects the joint influences of the (temporally constrained) semantic activation coming from the bottom-up volley of processing initiated by the stimulus, and the activation states previously induced by context information of various types (cf. Hagoort, 2005 for a different notion of “binding” in language processing; Baggio & Hagoort, in press further describe a framework emphasizing oscillatory interactions between frontal and temporal brain regions important for unifying bottom-up and top-down information in message-level formation). Federmeier & Laszlo lay out more detailed arguments, but the point germane to the present discussion is that their proposal – and the empirical literature that motivated it – implies that some aspects of context information may not become available before an initial semantic representation is formed, and thus may be reflected in processes subsequent to the N400.

Indeed, complementary to the evidence showing that contextual information seems to be used by the brain as early as possible to effect comprehension – in fact, in advance of any bottom-up stimulation when words are predictable (for review, see Federmeier, 2007) – are findings revealing that when relevant information becomes available later in the processing stream, the initial meaning representation may be revised or reshaped. In some cases, such as in figurative language or humor, reinterpretation may be the norm, as understanding a joke, metaphor, pun, or ironic statement often hinges on appreciating the multiple meanings of the contextual information, as cued by a critical word (such as the punchline of a joke). Thus, the contextual information on which a joke turns only becomes relevant once the disparity between the literal and humorous senses of the meanings emerges. In several studies of figurative language, Coulson and colleagues have observed a broad, slow, frontal/

anterior negativity, onsetting approximately 500 ms post-stimulus onset, that they have associated with a psychological mechanism called frame-shifting (Coulson & Kutas, 2001; Coulson & Lovett, 2004; Coulson & Williams, 2005). Because multiple meanings are often at play in non-literal language, context that has been understood with one interpretation may need to be revisited upon encountering the punchline of a joke or a pun; this reinterpretation of the contextual information shifts the "frame" created by initial reading of the context. As an example borrowed from Coulson and colleagues illustrates, a comprehender reading or listening to a sentence such as "By the time Mary had her fourteenth child, she had run out of names to call her…" would likely be led by the unfolding contextual information to build a message-level representation of a scenario involving the naming of a new baby. When the reader or listener then encounters the joke punch-word "husband", understanding the joke entails a shift in the frame set up by the context to a situation involving disdain for the spouse, and where "call" refers not to naming a baby, but to disparaging the husband.

Although figurative language (especially novel as opposed to conventionalized figurative language) might be particularly likely to rely upon semantic reanalysis processes such as frame shifting, Coulson and colleagues have argued and empirically shown that mechanisms involved in processing nonliteral language are more general, and that qualitatively similar processes are engaged during the comprehension of literal and nonliteral sentences (see, e.g., Coulson & Matlock, 2001; Coulson & Van Petten, 2002). Indeed, across several published studies (e.g. Kutas, 1993; Federmeier & Kutas, 2005; Davenport & Coulson, 2011), normal, plausible sentences that are at least moderately constraining seem to produce a late left-frontal negativity that appears similar to that detailed in studies of figurative language. To our knowledge, however, this effect has not been extensively analyzed or discussed in any of the studies of this type, and the processes that might be reflected by this effect have thus not been well-characterized for literal sentences. In most of these studies, only one or a few discrete levels of constraint were used, making it difficult to fully elucidate the possible functional significance of such an effect in literal sentences (or even to determine whether the effect is best characterized as a negativity or a positivity). Furthermore, most of the hypotheses of the prior studies were directed toward patterns of N400s, and as a consequence, less attention was paid to the later effects in these investigations.

Therefore, to probe later-stage processes that may be involved in semantic reanalysis in the context of literal language processing, in this study we examine ERP data recorded as participants read a set of entirely plausible and literal sentences that continuously varied in the amount of available contextual support. This parametric manipulation allows a precise characterization of the effects of sentential context on ERP responses -- such as the graded decrease of the N400 in association with increasing constraint -- and provides the opportunity to uncover responses that may be especially enhanced in a specific range of constraint, perhaps like the late negativity.

We propose that there may be a functional similarity between the late frontal negativity associated with frame shifting and the similar brain responses sometimes observed in ERP studies of literal sentence processing. In particular, in some instances, a sentence context may be compatible with multiple likely message-level representations. When the continuation of that context diverges from the message-level interpretation currently active for the comprehender, a process may be engaged to reconsider or reinterpret the contextual information, similar to the processes of the frame-shifting operation for comprehension of the humorous aspects of a joke.

Rather than attempt to construct sentences that may or may not promote frame-shifting like processes in literal sentences, the approach of this experiment is to use a set of normal plausible sentences with natural variation in the strength or likelihood of any particular message-level representation(s). This design offers the advantage of not drawing participants' attention to any specific (artificially contrived) aspect of sentence characteristics and guards against the possibility that participants would invoke an unusual comprehension strategy due to the experimental context. Because the materials were extensively normed, we are able to determine on an item-level basis the extent to which there may be a likely alternative interpretation for each sentence context. To tap the likelihood of a sentence engendering multiple interpretations, we employ a correlational approach at the item level.

Moderately constraining sentences are good candidates for eliciting effects related to reinterpretation because, unlike weakly constraining sentences, they are likely to engender the creation of robust, specific message-level representations as the sentence unfolds. However, unlike strongly constraining sentences, moderately constraining sentences are those in which multiple continuations are reasonably likely across people. Thus, a subset of people may initially interpret the sentence in a manner different from the meaning implied by the ending that is actually presented – requiring reinterpretation once the ending is encountered. For example, the sentence frame "The fertilizer enriched the…" is completed by most people with the word "soil" (79.4% in our sample). However, in being moderately rather than strongly constraining, these sentences also allow some variability, such that in some cases there is an alternative ending that a sizeable minority of participants generate instead. In this example, some people choose to complete the sentence with the word "grass" (8.8% in our sample). Thus, much of the time, the context leads the reader to construct a message-level interpretation wherein the purpose of fertilizer is to help create better soil for growing. However, for a subset of participants, the sentence instead evokes a scenario for lawn care in which fertilizer is used to improve the grass grown in the soil. For those latter participants, when presented with the most common completion, "soil", aspects of the message-level meaning are likely to be revised or revisited to accommodate or appreciate the new information -- in a manner perhaps similar to frame-shifting.

Not all moderately constraining sentences, however, will engender robust alternative interpretations in this way. Some, instead, may be moderately constraining because at least a subset of people have difficulty extracting a robust message-level representation from them at all. Such sentences are likely to lead to many, inconsistent alternate endings in the norming data (more similar to the pattern seen for weakly constraining contexts). For example, the sentence frame "At first the woman refused, but eventually she changed her…" is completed by the word "mind" by 85.3% of our norming participants. However, in this case, there is no strong alternate competitor, and there is more inconsistency in the responses given across participants (*answer*, *approach*, *attitude*, *behavior*, *clothes*, *decision*, *manner*, *name*, *view*). Thus, there is no clear likely alternative interpretation.

To operationalize the likelihood of a sentence context engendering a robust alternative interpretation -- and thus being a good candidate for eliciting processes involved in the reconsideration of that interpretation upon encountering the final word -- we counted the number of unique sentence endings in our norming data (cf. Dimigen et al., 2011). That is, those sentences that yield a single, or only a few, alternatives to the best completion are likely to be those that have a relatively consistent alternative reading. In contrast, sentence contexts that yield a large number of alternative completions (each given by one or only a couple of participants in the norming study) might be those for which participants are simply unable to derive a strong message-level meaning, rendering the sentences more similar to weakly constraining contexts.

Thus, we expect that our set of plausible literal sentences will engender the late frontal negativity in moderately to strongly constraining contexts, as observed in prior studies with similar materials (e.g. Kutas, 1993; Federmeier & Kutas, 2005). The continuous manipulation across the full range of constraint will allow a more precise characterization of the ranges in which the negativity is most prominently observed. Then, in order to better characterize the processes reflected by the effect, we examine possible explanations for the late frontal negativity with an items-level approach to analyzing the ERPs (cf. Laszlo & Federmeier, 2011), using the number of alternate completions from the norming data as a predictor of brain activity to determine whether the likelihood of a robust alternative interpretation modulates the size of the effect.

2 Method

2.1 Participants

Sixteen native English speakers (eight female, eight male) who reported no exposure to other languages before age 5 were included in the experiment (mean age 19.7, range 18–27). Participants were right-handed by self-report and as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), with a mean handedness score of .74 (four participants reported at least one left-handed immediate family member). All participants reported normal or corrected-to-normal vision, and none reported a history of neurological or psychiatric disorder or recent use of any psychoactive medications. Written informed consent was obtained from all participants, and experimental protocols were approved by the Institutional Review Board at the University of Illinois at Urbana-Champaign. Participants received course credit or monetary compensation for their time.

2.2 Materials

The experimental stimuli consisted of 300 entirely plausible sentences varying continuously through the full range of cloze probability (range of 0–100%)1. Sentences were normed for cloze probability at the University of California, San Diego, as previously described in published reports (Wlotko & Federmeier, 2007 and Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). Fifty sentences, taken from these prior studies, contained plausible but wholly unexpected endings (sometimes called "low-cloze" endings to differentiate them from anomalous endings; we refer to them here as "unexpected" as we do in our prior studies). These sentence frames were always weakly constraining (i.e. the cloze value of the word with the highest cloze probability for that frame did not exceed 42%), but the sentence-final unexpected endings were never the word with the highest cloze probability for the sentence frame. The actual cloze value for the unexpected words was near zero and did not exceed 10%. For the remaining 250 sentences, the sentence-final critical word was always the best completion (the word with the highest cloze probability for that frame). These sentences were divided into five cloze probability ranges with 50 sentences per condition. This constituted ranges of 10–30%, 30–50%, 50–75%, 75–90%, and 90–100%. Examples are presented in Table 1.

The sentence-final critical words in these cloze ranges did not differ in lexical characteristics of word frequency (mean=114, SD=3.7), length (mean=5.0, SD=0.3), concreteness (mean=520, SD=16.0), imagability (mean=543, SD=11.9), familiarity (mean=575, SD=4.1), or word class (80% nouns); all characteristics were retrieved from the MRC Psycholinguistic Database (Wilson, 1988). Mean sentence length across the six conditions was 10 words (SD=0.2). Because the characteristics were relatively homogenous, other

[1]Cloze probability (Taylor, 1953) is an empirically determined measure of sentential constraint, estimated by asking participants (34–35 participants for each item, in this case) to complete sentence frames with the word that first comes to mind and tallying the proportion of responses for each frame. We follow the prior literature in using cloze probability as an operationally defined measure of the amount of contextual information in sentence contexts.

groupings of cloze probability did not cause major differences in lexical characteristics. In addition, cloze probability is not correlated with any of the lexical characteristics. The sentences were pseudorandomly assigned to two lists with different orders of presentation, and each list was presented to half of the participants.

For each item in the stimulus set, the number of unique alternative completions generated by participants in the cloze norming was tabulated. For cloze probability calculations, synonyms and alternate forms were collapsed (e.g. boat/ship or phone/phones), and summed over both the first and two secondary completions generated by participants in the norming. To count uniquely generated sentence completions, each unique lexical item was counted separately, for the first generated cloze response only.

2.3 Procedure

Participants were seated in a dimly lit room 100 cm in front of a 21” CRT computer monitor. Each trial began with a warning sign (several pluses in the center of the screen) presented for 500 ms; the duration of the blank screen between the warning sign and the first word of the trial varied randomly from 500 to 1200 ms to avoid averaging in slow potentials associated with anticipation of sentence onset. Each word appeared in a Helvetica 22-point font with black text on a white background. Sentences were presented word-by-word in the center of the screen, and all words were presented for 200 ms with a 300 ms interstimulus interval. A 2.5 second pause followed each trial, which was then followed by a display of the prompt "Please go on when you are ready." Participants initiated the next trial with a button press. Participants were asked to minimize blinks, eye movements, and muscle activity while reading. They were instructed to read the sentences for comprehension while keeping in mind that they would be asked questions about what they had read at the conclusion of the experiment. The recording session began with a short set of practice sentences to acclimate the participants to the task situation. The main experimental session was divided into ten blocks of sentences that lasted 4–5 minutes each, with participants taking a short rest between each block.

After the recording session, participants completed a recognition test of the sentence materials, for which mean d' was 2.14, indicating participants were attending to the stimuli.

2.4 EEG Recording and Processing

EEG was recorded from 26 geodesically arranged sites on the scalp using silver-silver chloride electrodes embedded in an Electro-cap. The position of the cap was determined by placing the Midline Prefrontal electrode (corresponding to Fpz) at 10% of the nasion-inion distance from the nasion, the Midline Occipital electrode (corresponding to Oz) at approximately 10% of the distance from the inion, and the Midline Central (corresponding to Cz) electrode halfway between the mastoid processes. The electrodes were referenced online to the left mastoid and later referenced offline to the average of the left and right mastoids. Eye movements were monitored using a bipolar recording of the electro-oculogram (EOG) with electrodes placed on the outer canthus of each eye. Blinks were monitored with an electrode placed over the infraorbital ridge of the left eye, referenced to the left mastoid. Electrode impedances were kept below 4 kΩ and signals were amplified with Sensorium amplifiers set at a bandpass of 0.02 to 100 Hz. EEG was sampled at 250 Hz and saved on a hard drive. EEG records were examined and marked for EOG, electromyographic (EMG), or other artifactual contamination. Trials containing eye blinks were corrected (see Dale, 1994 for the procedure) and added back into the EEG record, except for participants with too few blink artifacts for effective correction (eight participants). The remaining artifactual trials were excluded from further analysis (13% overall; proportions rejected per bin, in order of decreasing cloze probability: 12%, 14%, 12%, 14%, 14%, 14%). ERPs were computed from 100 ms before the onset of critical words to 920 ms after. The data were rereferenced to the algebraic mean of the left and right mastoids, and averages of artifact-free ERPs were calculated for critical words, after subtraction of the 100 millisecond pre-stimulus baseline. Measurements were taken after a digital bandpass filter of .2 to 20 Hz was applied. As can be seen in Figure 1, all conditions elicited the typical ERP response to visual words including, at posterior sites, a positivity (P1 or P100) peaking between 100 and 150 ms, a negativity (N1 or N150) peaking between 150 and 200 ms, and a positivity (P2 or P200) peaking around 250 ms, and, at anterior sites, a negativity (N1) peaking around 100 ms and a positivity (P2) peaking around 230 ms.

3 Results

3.1 N400

To assess the relationship between cloze probability and N400 amplitude, we divided each of our 50-item bins in half, creating twelve 25-item bins, allowing more variability to be explained by cloze probability. We correlated average cloze probability and N400 mean amplitude (measured over a 280–480 ms window, centered around the peak of the difference between the 90–100% and Unexpected bins) at the group level for these 12 datapoints for each electrode site. Correlations (all statistically significant) ranged from .3 at the front of the head to .9 over posterior regions, increasing posteriorly and to the right, consistent with the canonical N400 distribution. A topographic map of interpolated (with spherical splines) correlation coefficients is shown in Figure 2. This result corroborates the many studies showing the close association of the amount of sentential context with N400 amplitude, validating our stimulus materials for the investigation of the use of sentential context in normal plausible sentences.

3.2 Late left-lateralized fronto-temporal negativity

Visual inspection of the grand averaged ERPs revealed a left-lateralized fronto-temporal negativity beginning 400–500 ms post-stimulus onset, largest for items in the 75–90% cloze range (see Figure 1). An omnibus repeated measures ANOVA of mean amplitudes in the 600–900 ms time window revealed a main effect of Condition [$F(5,75)=2.81$, $p=.022$, $\varepsilon=.9562$] and a Condition × Electrode interaction [$F(125,1875)=1.77$, $p=.05$, $\varepsilon=.105$]. To further characterize this effect, a difference wave was created by subtracting the ERP for Unexpected endings from the ERP for the 75–90% bin. A topographic map of this voltage difference measured between 600–900 ms post-stimulus onset is presented in Figure 2. The mean amplitude in this time window (measured at three left anterior/central sites at the focal point of the effect) for the 75–90% bin is significantly different from all other bins (all Fs, with 1 and 15 degrees of freedom, are greater than 6.0, leading to all $p<.0254$).

To our knowledge, this effect has not been well-characterized in the literature for normal plausible sentences of this type. As such, we used an items-level approach to determine what factors may be related to the left frontal negativity. That is, rather than creating single-subject ERPs for conditions of interest, we averaged together the evoked responses (after discarding artifactual trials) across subjects for each item (cf. Laszlo & Federmeier, 2011; Rey et al., 2009) and correlated these item amplitudes with the norming data.

While not significantly related to the magnitude of negativity (relative to baseline) across items in the whole stimulus set ($r=.06$, $p=.31$), the number of alternative endings correlates significantly with item-level negativity (measured over the same three left anterior channels

[2]For tests with more than 1 degree of freedom in the numerator, p-values are adjusted with the Huynh-Feldt correction for violation of sphericity, and the epsilon value is reported.

as above, from 600–900 ms) in the 75–90% cloze range ($r=.27$, $p=.05$), and this correlation is specific to this cloze range (correlations for the other ranges are between −.14 and .09 with all p-values > .32). A scatterplot of number of endings versus mean amplitude between 600–900 ms is displayed in Figure 3. As revealed by this plot, the fewer alternative completions there are, the larger the negativity for moderately strong contexts (i.e. more alternative completions leads to more positive amplitudes).

If the correlation analysis reveals something about the functional significance of the late negativity and is not due to chance, e.g. as a result of the particular channels we analyzed, then the pattern of correlation across the scalp should be similar to the distribution of the ERP effect. Figure 4 displays interpolated (with spherical splines) maps of item-level correlation coefficients over the scalp. Although the pattern of correlation over the scalp extends more posteriorly than the voltage difference between the moderately constrained items and the unexpected items, overall, the correlation map reveals a left frontal focus like the effect observed in the ERPs.

Further, ERPs were re-averaged to confirm that the effect of few versus many alternative completions revealed by the correlation analysis comprises an ERP difference similar in timing and morphology to the overall effect of the late frontal negativity. For the moderately constraining items (75–90% cloze), the bin was divided into sentences with fewer than 5 alternative completions, and sentences with 5 or more alternates. Figure 5 shows these averages at representative electrode sites over the head, along with the ERPs for all sentences from the 90–100%, 50–75% and Unexpected ranges for comparison. The averages are plotted from 700 ms prior to 1000 ms after sentence-final word onset (baseline of 100 ms prior to onset of penultimate word) to ensure the effects observed in sentence-final ERPs were not solely due to pre-stimulus brain activity.

Although a statistical test of few versus many completions within the moderately strong range of constraint would not be very informative given the already reported correlation within this constraint range, we conducted a confirmatory ANOVA to statistically demonstrate a difference in the effect of number of alternate completions in strong versus weak contexts (at the subject level). For this test, the Stochastic Optimization of Stimuli algorithm (SOS; Armstrong, Watson, & Plaut, 2012) was used to select 2 groups of 15 stimuli above 70% cloze (high constraint) and 2 groups of 15 stimuli between 20 and 50% cloze (low constraint) that differed as much as possible in the number of alternative endings but were as similar as possible in cloze probability (within each level of constraint). The minimum number of alternative completions was 2 (otherwise, all the "few completion" sentences would come from the 90–100% cloze range, which does not generate the negativity). For high constraint, this procedure resulted in a "few completions" bin that averaged 83% cloze and 3.8 alternate completions and a "many completions" bin that averaged 77% cloze and 7.9 alternate completions. For low constraint, the "few completions" bin averaged 36% cloze and 8.2 alternate completions and the "many completions" bin averaged 31% cloze and 19.2 alternate completions. A repeated measures ANOVA (by subjects) was conducted on mean ERP amplitudes for these bins, measured from 600–900 ms post-stimulus onset. As in the omnibus ANOVA above, all electrode sites were used so that the analysis was not based on the exact same channels as used in the correlation analysis. The results of this analysis indicated no main effect of Constraint [$F(1,15)=0.85$], no main effect of Number of Alternates [$F(1,15)=0.44$], but an interaction between these factors [$F(1,15)=4.77$, $p=.04$]. The pattern of means is as expected based on the correlation analysis: High Constraint, Few Endings: 0.98 μV (SEM=0.13); High Constraint, Many Endings: 2.06 μV (SEM=0.14); Low Constraint, Few Endings: 2.24 μV (SEM=0.16); Low Constraint, Many Endings: 1.61 μV (SEM=0.17).

3.3 Validation

To validate the relationship we uncovered between the size of the frontal negativity and the number of alternative endings generated in the cloze task, we reanalyzed the data from a prior published study that also manipulated cloze probability (among other factors; Federmeier et al., 2007). This dataset provides a chance to replicate the correlation in a completely independent sample, using a similar but not identical paradigm. Although not readily apparent in the grand average, a subset of the subjects (~ 31%) showed a negativity for moderately to strongly expected items, similar in timecourse, scalp distribution, and morphology to the negativity observed in this study. For these participants[3], we again found a significant correlation (r=.24; p=.05) between the magnitude of frontal negativity (again measured over the same three left anterior channels at the focal point of the effect, from 600–900 ms) and the number of alternative endings in a cloze range similar to the present study (70–90%)[4]. The p-value for the two results combined, given by $\chi^2=-2\Sigma[\ln(p)]$ (Fisher's combined probability test, see, e.g., Elston, 1991), with degrees of freedom equal to 2 times the number of independent p-values (2 in this case), is 0.018 (χ^2 =11.98, df=4). Thus, it is unlikely that the correlation in the current study was due to Type I error.

Not all participants in either the current study or in the re-analyzed validation study show the negative-going frontal effect. If the correlation of ERP amplitude and number of alternates is functionally related to the negativity effect, then it should not be observed for those participants who do not show the negativity. For the participants in Federmeier et al. (2007) who do not exhibit the negativity, the correlation with alternative endings is not significant (r=.11, p=.32). In other studies with comparable late frontal negative effects, similar individual variability has been reported. For example, in Coulson's studies of joke comprehension, the negativity is modulated by the ability to appropriately comprehend the jokes, handedness, and other factors (see, e.g., Coulson et al., 2001; 2004).

4 Discussion

The oft-cited "sentence paradox" – that the meaning of a sentence is determined by the individual words making up that sentence, yet the sentence can also shape the meaning of its words – lies at the nexus of many lines of research in the cognitive science of language. Studies querying when and how contextual information can affect the stimulus-driven processing of words have revealed multiple kinds of context effects that emerge at different times (and different time scales) and often in parallel (e.g., Hagoort & Van Berkum, 2007; Federmeier, 2007; Kuperberg, 2007; cf. Kutas, 2006). In this study, we further examined how word and context information interact over time, with a focus on processing similarities and differences across contexts that provide varying levels of constraint.

As expected, N400 amplitudes closely tracked variation in the predictability of sentence-final words, corroborating many prior findings (e.g., Kutas & Hillyard, 1984 and DeLong et al., 2005). In addition, we observed a later negativity with a left fronto-temporal scalp distribution that was most evidently enhanced for sentence completions of moderately strong constraint. That the simple manipulation of amount of sentential context resulted in

[3]A participant was defined as exhibiting the negativity if the mean amplitude of the ERP to the moderately/strongly expected items from that study was more negative than that to the weakly constrained unexpected items (a similar comparison to the present study) in the 11 anterior channels between 600–900 ms. Note that the negativity itself should be smaller in the prior study because the "expected" condition encompasses a broader cloze range than elicits the negativity in the current study (i.e. up to 100% cloze). Thus, it is not surprising that the negativity was not as evident in the prior study. To be conservative, the individual subject effects were based on the ERPs for the entire condition in the prior study. However, to align the correlation analyses across the two studies, items above 90% cloze were removed in the items analysis.

[4]The size of the correlation was similar when all participants from the prior study were included in this analysis, but did not reach statistical significance (r=.21, p=.07).

temporally and qualitatively distinct ERP patterns provides further support for the idea that comprehension processes are multiple in nature, and extended over time.

We found that the size of the late left-lateralized frontal negativity was (inversely) correlated at the item level with the number of unique alternative completions generated by participants in the cloze norming task. Within the moderately strong range of constraint, items with fewer unique completions generated in the offline cloze task elicited a larger (frontal) relative negativity. Further, this result was replicated in a separate study with a different group of participants, with substantially overlapping experimental materials. Of course, the number of alternative endings may be a proxy for some other aspect of the sentence contexts that drives the brain response[5]. However, the specificity of the correlation to the 75–90% cloze range points to hypotheses about the types of processes that could be reflected by the negativity.

In particular, based on similar findings in the ERP literature on nonliteral language processing (e.g., Coulson, 2001; 2004; 2005), we suggest that this response may be related to the reinterpretation of contextual information when participants had initially built a message-level representation consistent with an alternative reading of the sentence. For example, given the sentence frame “Each night the campers built a huge…”, the best completion “fire” is given by about 79% of participants. However, a substantial minority completed the sentence with the word “tent” instead, suggesting that the context had brought to mind for them a different type of building event. For participants who initially interpreted the context information in a manner consistent with the alternative completion, integrating the presented best completion into their mental representation may have required some reinterpretation of what they had just read. We suggest this could be similar to or have mechanisms in common with what Coulson has described as frame-shifting, a psychological process in which elements of a mental representation are reanalyzed in light of new incoming information. Indeed, Coulson reports a left frontal negativity in ERP studies of joke comprehension, interpreted as reflecting frame-shifting (Coulson et al., 2001; 2004; 2005), and further, in an eye-tracking study using joke materials, Coulson and colleagues have shown that participants literally look back while reading to reconsider contextual information under conditions for which frame-shifting is likely (Coulson, Urbach, & Kutas, 2006).

Some prior ERP investigations of sentence processing have revealed an ERP effect similar in polarity, timecourse, and distribution to the frame-shifting effects, but in plausible literal sentences (see Kutas, 1993; Federmeier & Kutas, 2005; Davenport & Coulson, 2011). Most of this work mainly examined N400 patterns and did not focus on the late effects. Further, these studies typically did not examine gradations of constraint, instead usually defining “strong” constraint to be above some threshold of cloze probability (often 50%, 60%, or 70% cloze). Our continuous variation of constraint allowed us to narrow the possibilities for the functional significance of such an effect in sentence comprehension. The moderately constraining contexts (the 75–90% range in our materials), especially those with few alternative completions, may be especially likely among plausible literal sentences to engender frame-shifting like processes. In order for such processes to occur, comprehenders need to be able to create a clear message-level interpretation of a context (an initial “frame”) – something that may be less likely for weakly constraining contexts and for some moderately constraining contexts, which will lead to high variability in the alternate responses for those items (e.g. "She wished she had brought something to….", completed by many different possibilities). Yet that interpretation must also ultimately be different from

[5]None of the stimulus characteristics we report for our materials were able to significantly explain variance in the late frontal negativity.

the one necessitated by the presented ending, which is unlikely to happen for very constraining contexts that typically permit only a single interpretation ("She had never seen a spider get tangled up in its own…", which elicits "web" from 97% of participants).

The frontal negativity differs from other late frontal effects that have been observed in sentence processing studies. In the context of strongly constraining sentence frames, Federmeier and colleagues (Federmeier et al., 2007, 2010) have reported a late positivity, largest over prefrontal electrode sites, in response to plausible but wholly unexpected completions. Federmeier et al. linked this effect to cases in which participants make strong predictions about upcoming words that are then disconfirmed. If our hypothesis of the late negativity – that it is elicited in cases wherein participants considered a different interpretation of the context than the one corresponding to the sentence-final word in the experiment – then why would these sentence-final critical words not be "unexpected" from the participants' point of view, and thus elicit the frontal positivity effect? We assert that the eliciting conditions for the two effects can be (at least partially) dissociated.

Interestingly, both types of brain responses seem to require that the ending be plausible, as these responses are generally not observed for semantic anomalies, which often generate a "post-N400" positivity, also known as the late positive complex (LPC), or sometimes P600, with a posterior scalp distribution (see Van Petten & Luka, 2006; van de Meerendonk et al., 2010). Thus, for both effects, a coherent message-level representation seems to be a necessary antecedent. However, the (pre)frontal positivity is elicited by wholly unexpected words (cloze probability near zero), whereas the negativity observed in the present study is elicited in response to the most common completion, albeit one that might entail a different reading of the context from that initially considered by a subset of the participants. Thus, N400 responses are large to the unexpected words that elicit the (pre)frontal positivity because these words are not good fits with the prior context. However, the moderately constrained expected endings that are associated with the frontal negativity elicit reduced N400 amplitudes, attesting to the fact that these words receive contextual support and are good, plausible endings that fit well with the contextual frame. Further, at the item level, N400 amplitude is not correlated with the size of the negativity[6] – i.e. it does not seem to be the case that items with the largest N400s, less facilitated by context, are the ones that elicit the largest negativity. Thus, the negativity-eliciting items may not signal an entirely incorrect prediction, but may initiate a reinterpretation of the contextual frame as we have described.

Further evidence that the (pre)frontal positivity is specifically linked to predictive processing comes from studies of age-related changes in comprehension. The positivity is absent, for example, in older adults, who are less likely or less able to engage predictive mechanisms (Federmeier et al., 2002, 2010; see also Wlotko et al., 2010). In contrast, we do not explicitly link the frontal negativity to predictive processing, as a contextual frame needs to be formed from the sentence context whether or not anticipatory processing has been engaged[7]. Indeed, a recent study examining age-related changes in the use of sentential contexts finds that the same older adults who fail to show evidence of predictive comprehension strategies do elicit the left-frontal negativity (as well as the correlation with

[6]Mean amplitudes for the N400 and late negativity were measured over spatially and temporally distant observations in an attempt to avoid component overlap: 330–430 ms at a right lateral occipital site for the N400 and 700–800 ms at a left lateral anterior/central site for the negativity. The correlation was not significant either for the entire set of items (r=.08, p=.17) or for the 75–90% cloze range (r=.17, p=.24).

[7]By "predictive" or "anticipatory" comprehension mechanisms, we specifically refer to pre-activation of information about likely upcoming stimuli based on the contextual information, as discussed in detail elsewhere (Federmeier & Kutas, 1999; Federmeier, 2007), and not to the building of a message-level representation or situation model that may eventually become incompatible with upcoming but not-yet-presented items.

number of alternate endings), perhaps even more robustly than younger adults do (Wlotko & Federmeier, in press). Accordingly, we link the frontal negativity to cases not in which predictions have been disconfirmed (by a very unexpected, somewhat unrelated item), but to cases in which the context needs to be reconsidered based on the current (plausible, contextually supported) input.

It is important to note that we do not take the presence of the frontal negativity as a sign that predictive comprehension mechanisms are absent. We assume that the healthy young adults in this study reading predictable sentence contexts engaged predictive mechanisms during comprehension as they have in our prior studies. However, our design does not include unexpected endings within strongly constraining contexts, thereby not allowing a chance to observe the consequences of disconfirmed predictions. It is also possible that these two types of mechanisms are not mutually exclusive. Indeed, one may ask why joke punch-words would not elicit a frontal positivity, as these words are by definition unexpected (see, e.g., Coulson & Kutas, 2001). One possibility is that the nature of the relationship between the context and the critical word determines which sets of processes are engaged (contextually supported in one of a set of possible interpretations of the context, requiring reinterpretation for the alternate readings; or somewhat unrelated to the context, requiring a "recovery" from an incorrect prediction). Alternatively, both types of processes could occur for one sentence, but on the average in one particular experimental condition, one of the processes is more likely or more robust, resulting in either a negative-going or positive-going ERP difference in that condition. Further, in tapping into the natural variation in the likelihood of multiple interpretations of a context, we are certainly faced with individual differences and idiosyncratic comprehension outcomes across trials and subjects. We suggest that the moderately constraining sentences provide the circumstances necessary to see the negativity in the ERP averages in that condition more robustly than in the lower constraint conditions[8], but it is quite possible that this process could be observed in more weakly constraining contexts under the right conditions (as in the studies of joke comprehension by Coulson and colleagues).

In some studies, left frontal negativities have been associated with working memory, particularly verbal working memory (e.g. Ruchkin et al., 1992), or with the retrieval of information from long term memory (e.g. Rosler, Heil, & Henninghausen, 1995); both types of effects often scale with the amount of mnemonic material (more negative for more items in memory). The effect observed here could reflect memory demands involved in building or maintaining multiple interpretations of the context, perhaps similar to maintenance or manipulation of verbal information due to syntactic or structural constraints (e.g. Kluender & Kutas, 1993). This could also be similar to frontal effects observed in response to referential and semantic ambiguity. Van Berkum and colleagues (see Van Berkum et al., 2007), for example, have described a sustained frontal negativity – what they call the Nref effect – linked to holding ambiguous referents (e.g., the pronoun "she") in mind until they can be resolved. Similarly, Lee and Federmeier (2009; cf. Hagoort & Brown, 1994) have described a sustained frontal negativity in response to semantically and syntactically ambiguous words (e.g., "duck") when syntactic constraints compel a particular reading, but semantic contextual constraints are not available to aid with selection of that meaning. In both of these cases, similar in some ways to the eliciting conditions for the negativity observed in this study, multiple representations are active in parallel, perhaps requiring selection and/or inhibition of one of the meanings.

[8]In these materials, only 5 items in the 30–75% cloze range have fewer than 5 alternative completions and no items under 30% cloze have fewer than 7 alternative completions; thus, only items in the 75–90% range are likely to have a consistent alternative interpretation.

Although the scalp distribution of an ERP effect cannot be used to directly infer its neural source, some have hypothesized the left inferior frontal cortex as a likely source for these effects on the basis of the expected distribution of frontal lobe sources and a consideration of other neuroimaging studies (e.g., Coulson & Lovett, 2004; Lee and Federmeier, 2009). Left frontal cortex is heterogeneous, even within classical language areas (i.e. Chein et al., 2002; Dapretto & Bookheimer, 1999; Friederici et al., 2000), and divisions have been proposed along a number of different dimensions, making it difficult to pinpoint exactly which frontal areas might be involved in generating the effect. However, one framework seems particularly compelling for our findings. Badre & Wagner (2007) describe a division in ventrolateral prefrontal cortex (VLPFC) wherein the anterior area is involved in the controlled retrieval of information from semantic memory, and the mid-posterior area is involved in selection among competing alternatives. The framework maps on to our study in that selection among competing alternative representations may be required in the reinterpretation of the context. This logic could be extended to all of the studies discussed above: selection of the appropriate (or inhibition of the inappropriate) representation may be needed for comprehending a joke, mapping referents, or getting to a single meaning of an ambiguous word.

Different areas of frontal cortex are likely to produce similar scalp potentials due to proximity of location. Thus, from ERP data alone, we cannot say whether our effect is generated by VLPFC, as predicted by Badre's conceptualization, or from more posterior and dorsal areas associated with holding items in memory (see, e.g., Smith et al., 1998), or some other brain area. Indeed, Van Berkum and colleagues conducted an fMRI study of co-reference, similar to their prior ERP work showing the Nref effect, and did not find strong evidence that the effect arises from left inferior frontal areas typically associated with verbal working memory. Instead, they associated the effect with activity in medial prefrontal areas commonly observed in conjunction with comprehending narratives or forming inferences (Nieuwland et al., 2007). In spite of their fMRI findings, however, an ERP study showed that the Nref is strikingly modulated by working memory span, such that subjects with low span scores showed virtually no effect in the ERP (Nieuwland & Van Berkum, 2006). Furthermore, for conditions of ambiguous reference, *decreases* in BOLD activity were observed in LIFG (and in fact, bilaterally). Clearly, more work is needed to understand how frontally distributed ERP responses observed during sentence processing relate to one another and to fMRI activity increases and decreases in specific frontal cortical areas or networks.

Irrespective of the specific neural source for the left fronto-temporal negativity we observe in the present study, our findings add to a growing literature revealing how multiple neurocognitive functions are used over time during the processing of even relatively simple, plausible sentences. Federmeier & Laszlo (2009) discussed the remarkable temporal stability of the N400, a marker of brain activity intimately tied with the processing of (potentially) meaningful stimuli in the world. The phenomenon of a temporal "deadline" on the initial formation of a semantic representation helps to explain the reciprocal relationship between the meaning of a sentence and the individual words that compose it, and thus to demystify at least part of the paradox. The brain uses whatever information it can as early as it can to help solve the complexities associated with normal everyday comprehension. However, information – from stimulus and from context – continuously accrues, and some information may not become available until after the initial formation of a representation. For example, when predictions can be made, they influence even the perceptual processing of words (e.g., Wlotko & Federmeier, 2007) and help shape initial semantic representations as indexed by the N400 (e.g., Federmeier & Kutas, 1999). However, when predictions are wrong, the brain cannot fully know the status of the incorrect prediction until after the word has initially been evaluated semantically. Therefore, the consequences of incorrectly predicting upcoming words manifest after the N400, as observed by Federmeier et al. (2007; 2010). Similarly, in the context of the present study, the reader's brain cannot know in advance whether its interpretation of the sentence is an alternative one to that entailed by the actually presented completion until after it has initially assessed the meaning of that word. Thus, the left fronto-temporal negativity that we have linked to reinterpreting the context frame emerges after 400–500 ms post-stimulus onset. These findings highlight the importance of considering how qualitatively different cognitive and neural mechanisms bring about comprehension – even of literal sentences – depending on when information becomes available over time.

Acknowledgments

This research was supported by the McDonnell Foundation and NIH grant AG026308 to Kara Federmeier, and NIH grant F31-DC009781 to Edward Wlotko. Scott Schulfer, Paul Ravenna, Andrew Dignan, and Cassandra Sung provided assistance with data collection and analysis. Marta Kutas provided helpful comments on an earlier version of this manuscript.

References

Armstrong BC, Watson CE, Plaut DC. SOS! an algorithm and software for the stochastic optimization of stimuli. Behavior Research Methods. 2012 Epub ahead of print.

Badre D, Wagner AD. Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia. 2007; 45(13):2883–2901. [PubMed: 17675110]

Baggio G, Hagoort P. The balance between memory and unification in semantics: A dynamic account of the N400. Language and Cognitive Processes. (in press).

Chein JM, Fissell K, Jacobs S, Fiez JA. Functional heterogeneity within Broca's area during verbal working memory. Physiology & Behavior. 2002; 77(4–5):635–639. [PubMed: 12527011]

Clifton, C.; Staub, A.; Rayner, K. Eye movements: A window on mind and brain. Amsterdam, Netherlands: Elsevier; 2007. Eye movements in reading words and sentences; p. 341-371.

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.

Coulson S, Kutas M. Getting it: Human event-related brain response to jokes in good and poor comprehenders. Neuroscience Letters. 2001; 316(2):71–74. [PubMed: 11742718]

Coulson S, Lovett C. Handedness, hemispheric asymmetries, and joke comprehension. Cognitive Brain Research. 2004; 19(3):275–288. [PubMed: 15062865]

Coulson S, Matlock T. Metaphor and the space structuring model. Metaphor and Symbol. 2001; 16(3–4):295–316.

Coulson S, Urbach TP, Kutas M. Looking back: Joke comprehension and the space structuring model. Humor: International Journal of Humor Research. 2006; 19(3):229–250.

Coulson S, Van Petten C. Conceptual integration and metaphor: An event-related potential study. Memory & Cognition. 2002; 30(6):958–968.

Coulson S, Williams RF. Hemispheric asymmetries and joke comprehension. Neuropsychologia. 2005; 43(1):128–141. [PubMed: 15488912]

Dale, AM. Unpublished Dissertation. San Diego, La Jolla, CA: University of California; 1994. Source localization and spatial discriminant analysis of event-related potentials: Linear approaches.

Dapretto M, Bookheimer SY. Form and content: Dissociating syntax and semantics in sentence comprehension. Neuron. 1999; 24(2):427–432. [PubMed: 10571235]

Davenport T, Coulson S. Predictability and novelty in literal language comprehension: An ERP study. Brain Research. 2011; 1418:70–82. [PubMed: 21925647]

DeLong KA, Urbach TP, Kutas M. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nature Neuroscience. 2005; 8(8):1117–1121.

Dimigen O, Sommer W, Hohlfeld A, Jacobs AM, Kliegl R. Coregistration of eye movements and EEG in natural reading: Analyses and review. Journal of Experimental Psychology.General. 2011; 140(4):552–572. doi:10.1037/a0023885. [PubMed: 21744985]

Federmeier KD. Thinking ahead: The role and roots of prediction in language comprehension. Psychophysiology. 2007; 44(4):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.

Federmeier KD, Kutas M. Aging in context: Age-related changes in context use during language comprehension. Psychophysiology. 2005; 42(2):133–141. [PubMed: 15787850]

Federmeier KD, Kutas M, Schul R. Age-related and individual differences in the use of prediction during language comprehension. Brain and Language. 2010; 115(3):149–161. [PubMed: 20728207]

Federmeier, KD.; Laszlo, S. The psychology of learning and motivation. Vol. vol 51. San Diego, CA, US: Elsevier Academic Press; 2009. Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory; p. 1-44.

Federmeier KD, McLennan DB, De Ochoa E, Kutas M. The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. Psychophysiology. 2002; 39(2):133–146. [PubMed: 12212662]

Federmeier KD, Van Petten C, Schwartz TJ, Kutas M. Sounds, words, sentences: Age-related changes across levels of language processing. Psychology and Aging. 2003; 18(4):858–872. [PubMed: 14692871]

Federmeier KD, Wlotko EW, De Ochoa-Dewald E, Kutas M. Multiple effects of sentential constraint on word processing. Brain Research. 2007; 1146:75–84. [PubMed: 16901469]

Forster KI. Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology. 1981; 33A(4):465–495.

Friederici AD, Opitz B, von Cramon DY. Segregating semantic and syntactic aspects of processing in the human brain: An fMRI investigation of different word types. Cerebral Cortex. 2000; 10(7): 698–705. [PubMed: 10906316]

Hagoort P. On Broca, brain, and binding: A new framework. Trends in Cognitive Sciences. 2005; 9(9): 416–423. [PubMed: 16054419]

Hagoort, P.; Brown, C. Perspectives on sentence processing. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.; 1994. Brain responses to lexical ambiguity resolution and parsing; p. 45-80.

Hagoort P, Van Berkum JJ. Beyond the sentence given. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 2007; 362(1481):801–811.

Holcomb PJ, Grainger J. Exploring the temporal dynamics of visual word recognition in the masked repetition priming paradigm using event-related potentials. Brain Research. 2007; 1180:39–58. [PubMed: 17950262]

Kamide Y, Altmann GTM, Haywood SL. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. Journal of Memory and Language. 2003; 49(1):133–156.

Kluender R, Kutas M. Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. Journal of Cognitive Neuroscience. 1993; 5(2):196–214.

Kuperberg GR. Neural mechanisms of language comprehension: Challenges to syntax. Brain Research. 2007; 1146:23–49. [PubMed: 17400197]

Kutas M. Event-related brain potentials (ERPs) elicited during rapid serial visual presentation of congruous and incongruous sentences. Electroencephalography and Clinical Neurophysiology. 1987; 40(Supplement):406–411.

Kutas M. In the company of other words: Electrophysiological evidence for single-word and sentence context effects. Language and Cognitive Processes. 1993; 8(4):533–572.

Kutas M. One lesson learned: Frame language processing--literal and figurative--as a human brain function. Metaphor and Symbol. 2006; 21(4):285–325.

Kutas M, Federmeier KD. Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). Annual Review of Psychology. 2011; 62:621–647.

Kutas, M.; Federmeier, KD.; Staab, J.; Kluender, R. Language. In: Cacioppo, JT.; Tassinary, LG.; Berntson, GG., editors. Handbook of psychophysiology (3rd ed.). 3rd ed.. New York, NY, US: Cambridge University Press; 2007. p. 555-580.

Kutas M, Hillyard SA. Brain potentials during reading reflect word expectancy and semantic association. Nature. 1984; 307(5947):161–163. [PubMed: 6690995]

Laszlo S, Federmeier KD. A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. Journal of Memory and Language. 2009; 61(3): 326–338. [PubMed: 20161064]

Laszlo S, Federmeier KD. The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. Psychophysiology. 2011; 48(2): 176–186.

Lee C, Federmeier KD. Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. Journal of Memory and Language. 2009; 61(4): 538–555. [PubMed: 20161361]

McClelland JL, St. John M, Taraban R. Sentence comprehension: A parallel distributed processing approach. Language and Cognitive Processes. 1989; 4(3–4):287–335.

Nieuwland MS, Petersson KM, Van Berkum JJA. On sense and reference: Examining the functional neuroanatomy of referential processing. NeuroImage. 2007; 37(3):993–1004. [PubMed: 17611124]

Nieuwland MS, Van Berkum JJA. Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. Brain Research. 2006; 1118(1):155–167. [PubMed: 16956594]

Oldfield RC. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia. 1971; 9(1):97–113. [PubMed: 5146491]

Rey A, Dufau S, Massol S, Grainger J. Testing computational models of letter perception with item-level event-related potentials. Cognitive Neuropsychology. 2009; 26(1):7–22. [PubMed: 18608320]

Rosler F, Heil M, Hennighausen E. Distinct cortical activation patterns during long-term memory retrieval of verbal, spatial, and color information. Journal of Cognitive Neuroscience. 1995; 7(1): 51–65.

Schwanenflugel PJ, LaCount KL. Semantic relatedness and the scope of facilitation for upcoming words in sentences. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1988; 14(2):344–354.

Ruchkin DS, Johnson R, Grafman J, Canoune H. Distinctions and similarities among working memory processes: An event-related potential study. Cognitive Brain Research. 1992; 1(1):53–66. [PubMed: 15497435]

Simpson, GB., editor. Understanding word and sentence. Oxford, England: North-Holland; 1991.

Smith EE, Jonides J, Marshuetz C, Koeppe RA. Components of verbal working memory: Evidence from neuroimaging. Proceedings of the National Academy of Sciences of the United States of America. 1998; 95(3):876–882. [PubMed: 9448254]

Swinney DA. Lexical access during sentence comprehension: (re)consideration of context effects. Journal of Verbal Learning & Verbal Behavior. 1979; 18(6):645–659.

Taylor WL. Cloze procedure: A new tool for measuring readability. Journalism and Mass Communication Quarterly. 1953; 30(4):415–433.

Van Berkum JJA. Understanding sentences in context: What brain waves can tell us. Current Directions in Psychological Science. 2008; 17(6):376–380.

Van Berkum JJA, Koornneef AW, Otten M, Nieuwland MS. Establishing reference in language comprehension: An electrophysiological perspective. Brain Research. 2007; 1146:158–171. [PubMed: 16916496]

Van Berkum JJA, van den Brink D, Tesink CMJY, Kos M, Hagoort P. The neural integration of speaker and message. Journal of Cognitive Neuroscience. 2008; 20(4):580–591. [PubMed: 18052777]

Van Berkum JJA, Zwitserlood P, Hagoort P, Brown CM. When and how do listeners relate a sentence to the wider discourse? evidence from the N400 effect. Cognitive Brain Research. 2003; 17(3): 701–718. [PubMed: 14561457] van de Meerendonk N, Kolk HHJ, Vissers CTWM, Chwilla DJ. Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. Journal of Cognitive Neuroscience. 2010; 22(1):67–82. [PubMed: 19199401]

Van Petten C. A comparison of lexical and sentence-level context effects in event-related potentials. Language and Cognitive Processes. 1993; 8(4):485–531.

Van Petten C. Words and sentences: Event-related brain potential measures. Psychophysiology. 1995; 32(6):511–525. [PubMed: 8524986]

Van Petten C, Coulson S, Rubin S, Plante E, Parks M. Time course of word identification and semantic integration in spoken language. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1999; 25(2):394–417.

Van Petten C, Luka BJ. Neural localization of semantic context effects in electromagnetic and hemodynamic studies. Brain and Language. 2006; 97(3):279–293. [PubMed: 16343606]

Wilson M. MRC psycholinguistic database: Machine usable dictionary, version 2.00. Behavior Research Methods, Instruments, and Computers. 1988; 20(1):6–11.

Wlotko EW, Federmeier KD. Finding the right word: Hemispheric asymmetries in the use of sentence context information. Neuropsychologia. 2007; 45(13):3001–3014. [PubMed: 17659309]

Wlotko EW, Federmeier KD. Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. Psychophysiology. (in press).

Wlotko EW, Lee C, Federmeier KD. Language of the aging brain: Event-related potential studies of comprehension in older adults. Language and Linguistics Compass. 2010; 4(8):623–638. [PubMed: 20823949]

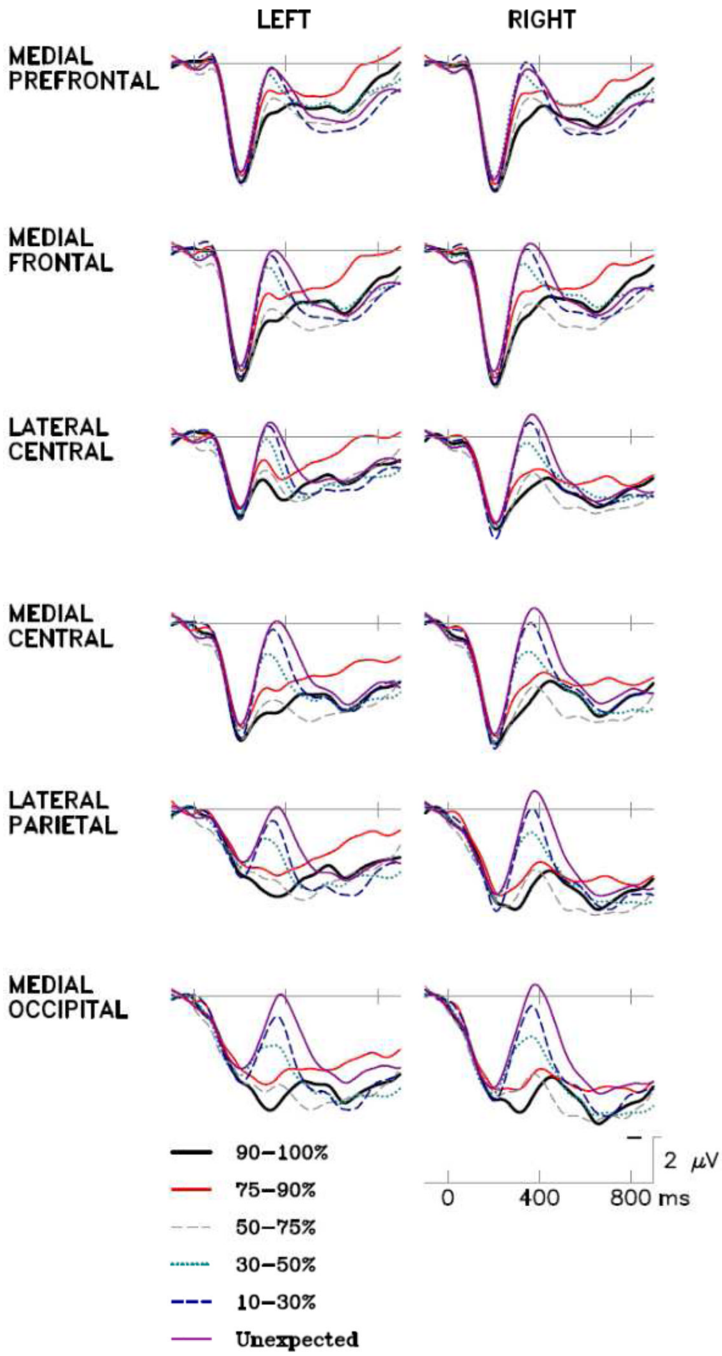

Figure 1.

Grand-average ERPs for the six 50-item cloze probability bins at 12 electrode sites. Whereas the N400 is graded with cloze probability, the later left fronto-temporal negativity is most enhanced for the 75–90% cloze range. Negative is plotted up.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

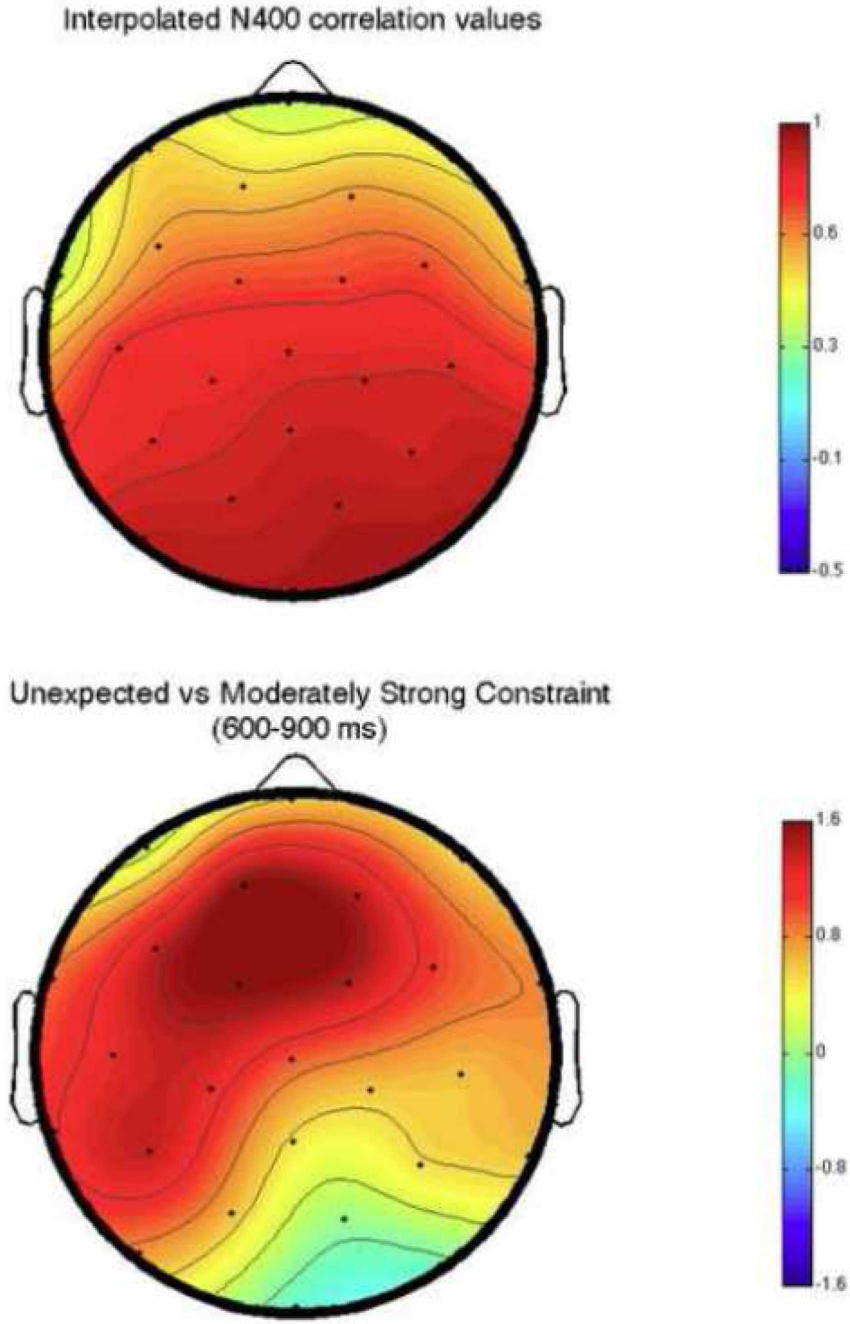

Figure 2.

Topographic maps of correlation coefficients for N400 amplitude vs cloze probability, and for the difference between Unexpected items and the 75–90% cloze bin averaged over 600–900 ms post-stimulus onset.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

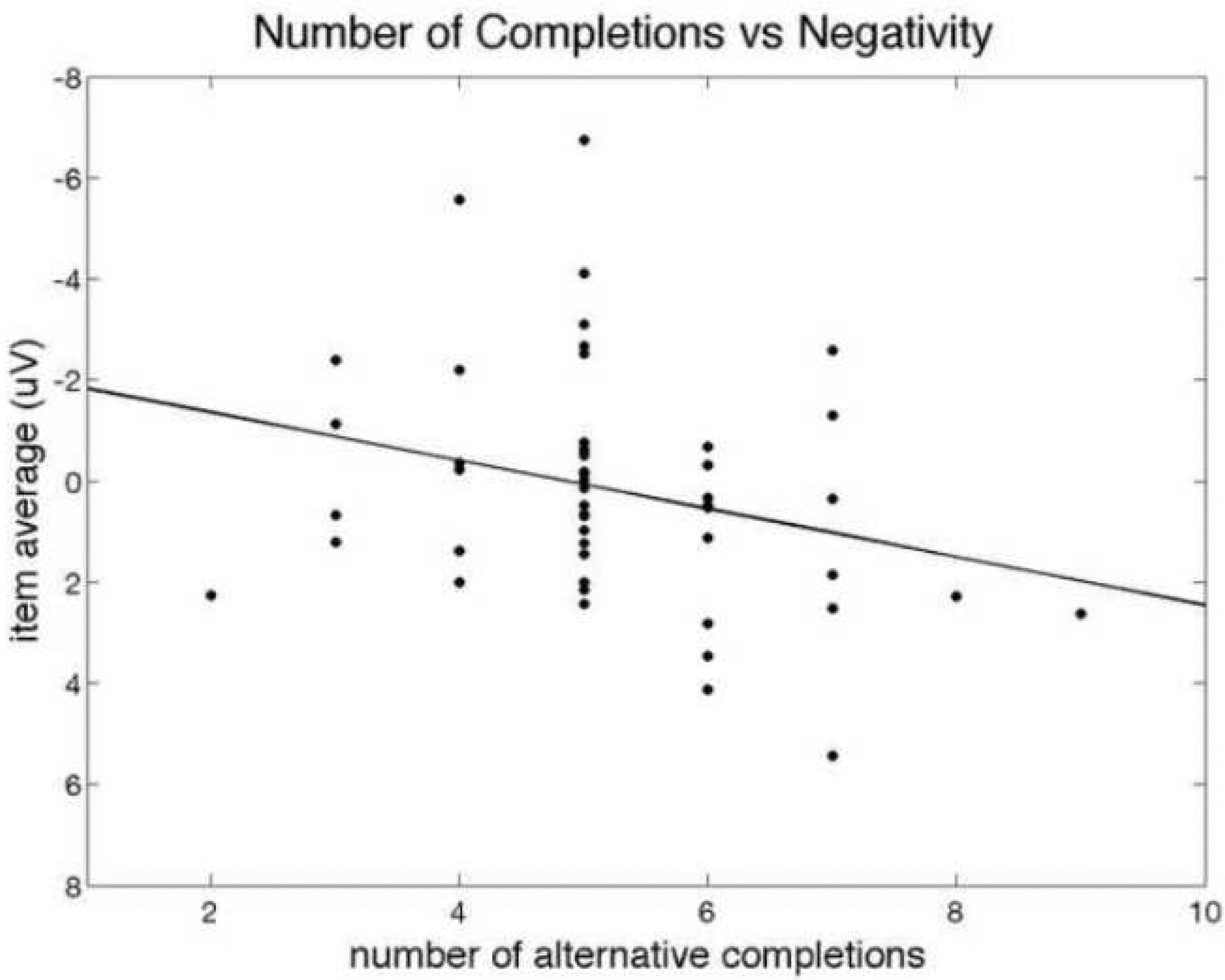

Figure 3.
Scatterplot showing the relationship between item-level negativity (600–900 ms post-stimulus onset at left frontal sites) and the number of unique alternate completions for the item's sentence context, for the 75–90% cloze range. Negative is plotted up.

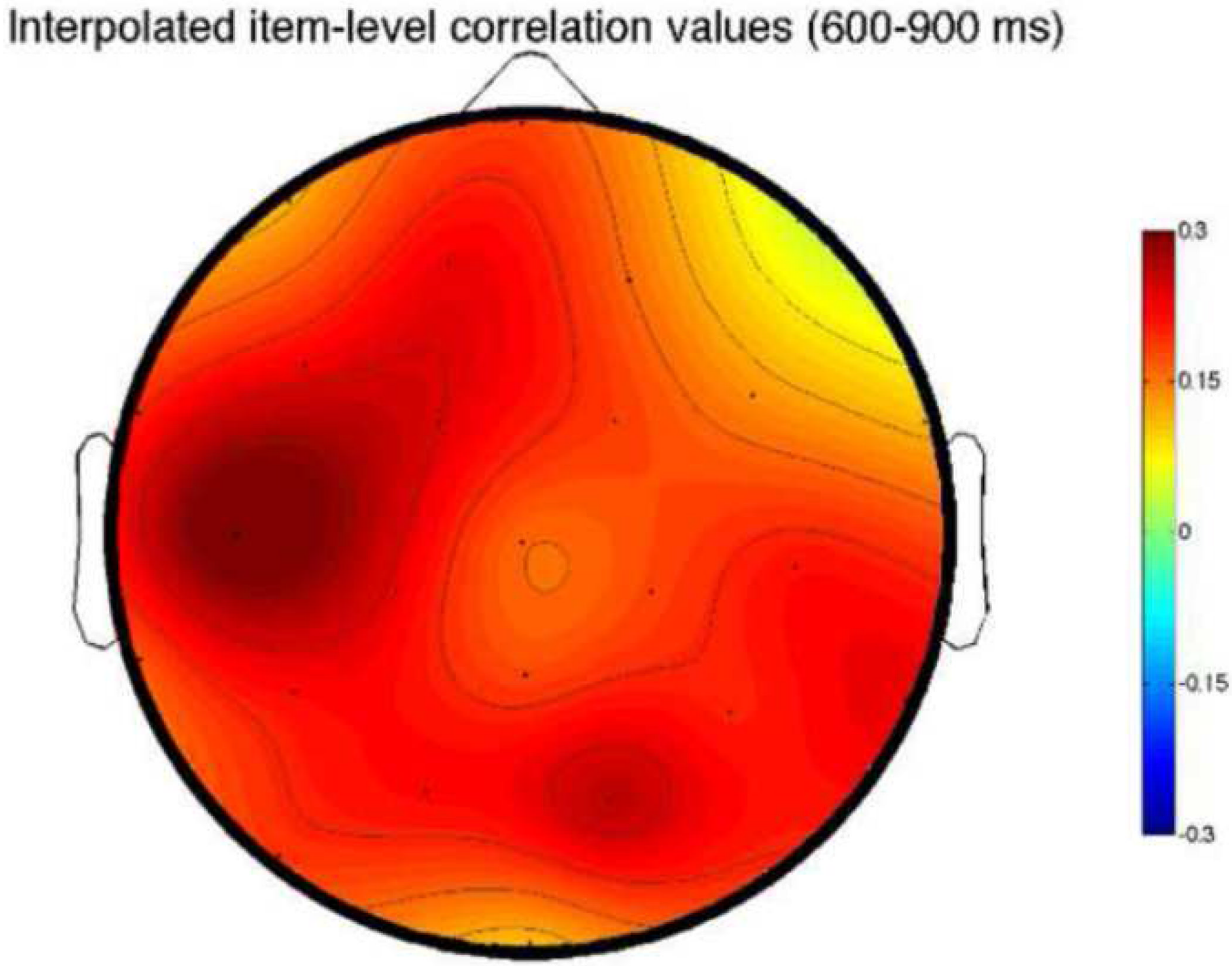

Figure 4.
Topographic map of correlation coefficients for the items analysis of mean amplitudes in the 600–900 ms timewindow and number of alternative completions for each sentence, within the moderately constraining range (75–90%). Like the scalp distribution of the overall ERP effect, the correlation map shows a left-frontal focus.

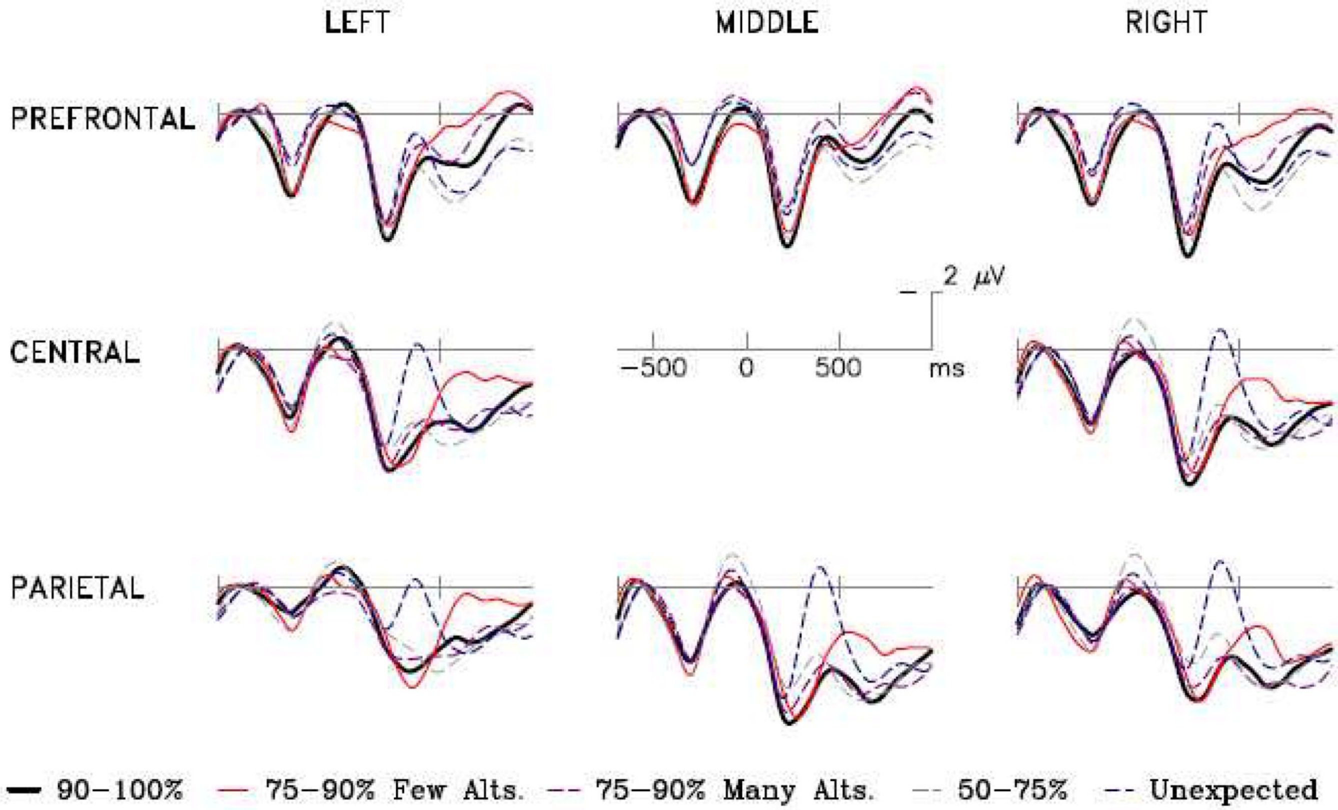

Figure 5.
ERPs for moderately constraining contexts with few and many alternative completions, plotted from 700 ms prior to 1000 ms after the onset of the sentence-final word (penultimate word onsets 500 ms prior to final word). Waveforms for the 50–75%, 90–100%, and Unexpected items are shown for reference. The effect of few versus many alternate completions is similar in timecourse and distribution to the overall effect of the late negativity.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Cloze Bin	Sentence	Best Completion (Cloze)	Unique Completions	Completions
90 – 100	The little girl refused to go to sleep until he told her a	story (100%)	1	*n/a*
	Tricia had never seen a spider get tangled up in its own	web (97.1%)	2	legs (2.9%)
75 – 90 (few alternates)	He was cold most of the night and finally got up to get another	blanket (85.3%)	4	drink (8.8%), shirt (2.9%), sweater (2.9%)
	After two hours of hard work they decided to take a short	break (77.1%)	4	nap (17.1%), cut (2.9%), walk (2.9%)
75 – 90 (more alternates)	Don't touch the wet	paint (85.3%)	6	bed (2.9%), cement (2.9%), dog (2.9%), outlet (2.9%), sand (2.9%)
	I just had a new sound system installed in my	car (79.4%)	7	house (5.9%), garage (2.9%), home (2.9%), ride (2.9%), room (2.9%), truck (2.9%)
50 – 75	She pulled her head out from under the faucet and reached for a	towel (70.6%)	11	beer, cup, drink, napkin, pill, purse, screwdriver, toothbrush, wrench, *no response*
	Jim was saving boxes for a friend who was	moving (64.7%)	11	busy, recycling, away, collecting, dead, desperate, drunk, gone, poor, sick
30 – 50	He was caught stealing a	car (47.1%)	17	purse, wallet, base, baseball, butterfinger, coat, diamond, dog, erasers, fish, gun, horse, knife, lemon, rug, shirt
	She wished she had brought something to	eat (34.3%)	10	read, do, give, show, wear, Canada, him, work, *no response*
10 – 30	It was time to hang the new	pictures (23.5%)	20	poster, banner, calendar, clock, clothes, coat, cord, decorations, drapes, frame, ornaments, painting, portrait, wreath, *etc.*
	They went to see the famous	actor (11.7%)	25	artist, building, elephant, gymnast, house, landmark, movie, magician, museum, painting, play, place, show, statue, *etc.*
Unexpected	They waited a long time to see the	grades		
	Rushing out he forgot to take his	camera		